

Neuroinformatics in Functional Neuroimaging

Finn Årup Nielsen
Informatics and Mathematical Modelling
Technical University of Denmark

2002-08-30

Abstract

This Ph.D. thesis proposes methods for information retrieval in functional neuroimaging through automatic computerized authority identification, and searching and cleaning in a neuroscience database.

Authorities are found through cocitation analysis of the citation pattern among scientific articles. Based on data from a single scientific journal it is shown that multivariate analyses are able to determine group structure that is interpretable as particular “known” subgroups in functional neuroimaging. Methods for text analysis are suggested that use a combination of content and links, in the form of the terms in scientific documents and scientific citations, respectively. These included *context sensitive author ranking* and automatic labeling of axes and groups in connection with multivariate analyses of link data.

Talairach foci from the BrainMap™ database are modeled with conditional probability density models useful for exploratory functional volumes modeling. A further application is shown with *conditional outlier detection* where abnormal entries in the BrainMap™ database are spotted using kernel density modeling and the redundancy between anatomical labels and spatial Talairach coordinates. This represents a combination of simple term and spatial modeling. The specific outliers that were found in the BrainMap™ database constituted among others: Entry errors, errors in the article and unusual terminology.

Statistical analysis and visualization have received much attention in neuroinformatics for functional neuroimaging and a large set of methods have been developed. Some of the most important analysis methods are reviewed with emphasis on cluster analysis, singular value decomposition, Molgedey-Schuster independent component analysis and linear models with FIR-filters. Furthermore, canonical ridge analysis is introduced as a mean for analysis of singular data. It can be viewed as a regularized canonical correlation analysis and in the limit of infinite regularization this is similar to a type of partial least squares. The model is also related to redundancy analysis, thus canonical ridge analysis subsumes different multivariate analyses and the solutions between them can be found by varying a continuous regularization parameter.

Scientific and information visualization methods are also reviewed with emphasis on VRML-based 3D visualization for functional neuroimaging results.

Dansk resumé

Denne Ph.D.-afhandling foreslår metoder til informationssøgning i forbindelse med funktionel hjernebilleddannelse gennem automatiseret og computerbaseret autoritetsbestemmelse og gennem søgning og rensning i en neurovidenskabelig database.

Autoriteter bliver fundet gennem kociteringsanalyse af citeringsmønsteret blandt videnskabelige artikler. Baseret på data fra et enkelt videnskabeligt tidsskrift bliver det vist at flerdimensionelle analysemetoder er i stand til at bestemme gruppestrukturer der er fortolkelige som visse “kendte” undergrupper i funktionel hjernebilleddannelse. Metoder til tekstanalyse er foreslået der bruger en kombination af indhold og netværksled, henholdsvis i form af termer i videnskabelige dokumenter og videnskabelige citeringer. Dette omfatter *sammenhængsfølsom forfatterarrangordning* og automatisk mærkning af akser og grupper i forbindelse med flerdimensionelle analyser af netværksdata.

Talairach punkter fra BrainMap™ databasen er modelleret med betinget tæthedsfordelingsmodeller anvendelige til udforskende funktionel volumemodellering. En anden anvendelse er vist med betinget udligger-detektion, hvor unormale registreringer i BrainMap™ databasen er opdaget ved hjælp af kernetæthedsmodellering og redundansen mellem anatomiske betegnelser og rumlige Talairach koordinater. Dette repræsenterer en kombination af simpel term og rumlig modellering. De specifikke udliggere der blev fundet i BrainMap™ databasen omfattede blandt andre: Registreringsfejl, fejl i artiklen og usædvanlig terminologi.

Statistisk analyse af visualisering har modtaget meget opmærksomhed indenfor neuroinformatik for funktionel hjernebilleddannelse og et stort sæt metoder er blevet udviklet. Nogle af de vigtigste analysemetoder er gennemgået med vægt på klusteranalyse, singular værdi-dekomposition, Molgedey-Schuster uafhængig komponentanalyse og lineære modeller med FIR-filtre. Endvidere er kanonisk *ridge* analyse introduceret som et middel til analyse af singulære data. Modellen kan betragtes som en regulariseret kanonisk korrelationsanalyse og i grænsen mod uendelig regularisering er den sammenfaldende med en type af *partial least squares*. Modellen er også relateret til redundansanalyse og indbefatter således forskellige flerdimensionelle analysemetoder, og løsninger mellem dem kan findes ved at variere en kontinuert regulariseringsparameter.

Videnskabelig og informations-visualiseringsmetoder er også gennemgået med vægt på VRML-baseret 3D visualisering af resultater fra funktionel hjernebilleddannelse.

This thesis serves as partial fulfillment of the requirements for the Ph.D. degree.

The work has been carried out at Informatics and Mathematical Modelling at the Technical University of Denmark with financial aid from the Danish Research Councils through THOR Center for Neuroinformatics and was supervised by Lars Kai Hansen and Jan Larsen.

Finn Årup Nielsen, Lyngby, Denmark, 2002–08–30

Contents

Abstract	3
Dansk resumé	5
Contents	7
List of Figures	11
List of Tables	15
1 Introduction	17
1.1 What is functional neuroimaging?	17
1.2 Why functional neuroimaging?	17
1.3 Computers, mathematics and statistics in functional neuroimaging	18
1.4 Contribution	18
1.5 Outline	19
2 Brain, mind and measurements	21
2.1 Behavioral and cognitive components	21
2.2 Specialization of the Brain	22
2.2.1 Microscopic structure	23
2.2.2 Functional specialization	24
2.3 Coupling	24
2.3.1 Stimulus-neuronal coupling	25
2.3.2 Hemodynamics	25
2.3.3 Deactivation	26
2.3.4 Coupling nonlinearity	26
2.4 Functional neuroimaging and other brain measurement techniques	27
2.4.1 Electrophysiology	27
2.4.2 EEG, MEG, EIT	27
2.4.3 Computerized tomography	28
2.4.4 Positron emission tomography and single photon emission computed tomography	28
2.4.5 Magnetic resonance imaging (MRI)	29
2.4.6 Optical methods: Optical intrinsic signal, near-infrared spectroscopy and voltage sensitive dyes	30
2.5 Stimulation of the brain	31
3 Analysis	33
3.1 Models	33
3.2 Estimation	34
3.2.1 Optimization	35
3.2.2 Regularization and priors	37
3.2.3 Non-orthogonality of design: Correlation between covariates	38
3.3 Testing	38
3.3.1 The number of samples and learning curve	39
3.3.2 Correlated and heterogeneous residuals	40
3.3.3 Simultaneous inference	40
3.4 Functional neuroimaging analysis	41
3.5 Preprocessing in functional neuroimaging	41
3.5.1 Reconstruction	42

3.5.2	Stripping	42
3.5.3	fMRI: Inhomogeneity correction, antialiasing, magnetization start-up	42
3.5.4	Motion correction	43
3.5.5	Coregistration — intermodality image registration	43
3.5.6	Spatial normalization	43
3.5.7	Segmentation	45
3.5.8	Confounds and nuisances in PET and fMRI: Removal of physiological noise,	45
3.5.9	Spatial Filtering	48
3.5.10	Reduction of the number of analyzed voxels	48
3.6	Analysis of functional neuroimages	48
3.7	Unsupervised methods	49
3.8	Cluster analysis	50
3.8.1	Choosing the number of groups	51
3.8.2	Spatial prior in clustering	52
3.8.3	Cluster analysis in functional neuroimaging	52
3.9	Principal component analysis and singular value decomposition	52
3.9.1	Probabilistic PCA	54
3.9.2	Factor analysis	54
3.9.3	Multidimensional scaling	54
3.9.4	SVD/PCA in functional neuroimaging	55
3.10	Non-negative matrix factorization	55
3.11	Independent Component Analysis	56
3.11.1	Molgedey-Schuster ICA	56
3.11.2	ICA in functional neuroimaging	57
3.11.3	Examples on Molgedey-Schuster ICA	57
3.12	Probability density estimation	60
3.12.1	Mixture models	60
3.12.2	Kernel methods	61
3.12.3	Probability density estimation in functional neuroimaging	62
3.13	Novelty and outlier detection	62
3.14	General model for unsupervised methods	63
3.15	Supervised modeling	64
3.16	Linear modeling	64
3.16.1	Regression	64
3.16.2	Time-series modeling of fMRI signals	66
3.16.3	Cross-correlation	67
3.16.4	Delay	67
3.16.5	Random effects	69
3.17	Nonlinear modeling	69
3.17.1	Artificial neural networks	70
3.17.2	Nonlinear modeling in functional neuroimaging	71
3.18	Canonical analyses	71
3.18.1	Canonical correlation analysis	71
3.18.2	Partial least squares	72
3.18.3	Orthonormalized PLS, redundancy analysis, etc.	74
3.18.4	Canonical ridge analysis	74
3.18.5	Canonical correlation analysis with singular matrices	75
3.18.6	Bilinear modeling	76
3.18.7	Optimization of the ridge parameter and subspace dimension	78
3.18.8	Generalization of canonical analysis	78
3.18.9	Canonical analyses in functional neuroimaging	79
3.18.10	Example on canonical analysis	80
3.19	From image to points	80
4	Visualization	83
4.1	Rendering techniques	83
4.1.1	3D polygon based visualization	83
4.1.2	Volume rendering	85
4.2	Visualization techniques	85
4.2.1	Polygon generation	85

4.2.2	Surface generation from contours	86
4.2.3	Glyphs	86
4.2.4	Network visualization	87
4.2.5	Stereovision	88
4.3	Scientific visualization in functional brain mapping	89
4.4	Information visualization for functional brain mapping	91
5	Neuroinformatics	93
5.1	Neuroinformatics	93
5.2	Neuroinformatics tools for functional neuroimaging	93
5.2.1	BrainMap™	94
5.2.2	Talairach Daemon	95
5.2.3	Connectivity database and analysis	95
5.2.4	Data centers	96
5.2.5	Library-like services	96
5.3	Meta-analysis	97
5.3.1	Meta-analysis in functional neuroimaging	97
5.4	Text analysis	99
5.5	Term analysis	100
5.5.1	Term identification — tokenization	100
5.5.2	Word stemming	101
5.5.3	Stop word and single instance words elimination	101
5.5.4	Term weighting	101
5.5.5	Other document elements	101
5.5.6	Analysis techniques for term matrices	102
5.6	Link analysis	102
5.6.1	Ranking	104
5.6.2	Categorization and clustering	105
5.6.3	Author cocitation analysis	105
5.7	Combining link and term analysis	107
5.7.1	Context sensitive author ranking	107
5.8	Example of analysis with “NeuroImage”	107
5.8.1	Author cocitation analysis	108
5.8.2	Coauthor analysis	112
5.8.3	Analysis of journals	114
5.8.4	Finding related authors	116
5.8.5	Context sensitive author ranking	118
5.8.6	Further discussion	119
5.9	Example of analyses with BrainMap™	120
5.9.1	Finding outliers	120
5.9.2	Modeling of the functional relationship — functional volumes modeling	126
5.9.3	Finding related experiments	127
6	Conclusion and discussion	131
A	Notation and terminology	133
A.1	Symbol list with main notations	133
A.1.1	Data Matrix	134
A.2	Word list	134
A.3	General Abbreviations	140
A.4	Anatomical names and abbreviations	142
B	Derivations	145
B.1	Principal component analysis as constrained variance maximization	145
B.2	Ridge regression and singular value decomposition	145
B.3	Mutual information	148
B.4	The Bilinear model	149
B.4.1	Canonical correlation analysis	149
B.5	Molgedey-Schuster ICA	149
B.6	Distance between two vectors	151
B.6.1	The product of two Gaussian distributions	151

B.6.2 Kullback-Leibler distance	151
C Example of web-services	153
D Acknowledgment	155
E Articles and abstracts	157
Bibliography	267
Author Index	319
Index	337

List of Figures

2.1	Segregation of mathematics cognitive components	22
2.2	Core processes involved in word production	22
2.3	Main components of NeuroNames brain Hierarchy	23
2.4	Coupling	24
2.5	Hemodynamic response function	26
2.6	Model of a PET scanner	29
3.1	Model with input and output.	33
3.2	Estimation chain	34
3.3	Normalized cumulated periodogram	47
3.4	Cluster types	50
3.5	Demonstration of Molgedey-Schuster ICA	58
3.6	Demonstration of Molgedey-Schuster ICA	59
3.7	Kernel density estimation	62
3.8	Resampling jittered events	68
3.9	Cross-correlation APCA network	73
3.10	Linear approximation APCA network	74
3.11	Canonical ridge analysis of an fMRI data set	80
4.1	Example on a VRML visualization	84
4.2	Sequence of volume renderings	85
4.3	Examples of 3D glyphs in Talairach space	86
4.4	Amber/blue anaglyph stereogram	89
4.5	Information visualization of a functional neuroimaging experiment	92
5.1	Organization of data in BrainMap™	95
5.2	A processing scheme for the vector space representation.	100
5.3	Graph in connection with author cocitation analysis	103
5.4	Cited authors and citing documents as nodes	103
5.5	Extraction of authors	108
5.6	Author cocitation analyses on cited author	110
5.7	Second and third eigenauthor	111
5.8	Coauthor bullseye plot	113
5.9	Coauthor bullseye plot	114
5.10	Journal cocitation analysis	117
5.11	Journal cocitation analysis	118
5.12	Processing scheme for finding outliers in BrainMap	121
5.13	Probability density estimation of “cerebellum”	122
5.14	Novelty table for locations	124
5.15	Probability density estimation of “lobe”	125
5.16	Corner Cube of “cerebellum” locations	126
5.17	Functional volumes modeling VRML screenshot	127
5.18	VRML screenshot of novelties of locations	128
5.19	A processing scheme for finding related BrainMap™ experiments based on volume comparisons.	128
5.20	Query volume and Results from a search	129
C.1	Web-services	154
E.1	HBM’97 abstract: FIR filter	159

E.2	HBM'99 abstract: Artificial neural network	160
E.3	IEEE TMI smooth FIR article. Page 1	161
E.4	IEEE TMI smooth FIR article. Page 2	162
E.5	IEEE TMI smooth FIR article. Page 3	163
E.6	IEEE TMI smooth FIR article. Page 4	164
E.7	IEEE TMI smooth FIR article. Page 5	165
E.8	IEEE TMI smooth FIR article. Page 6	166
E.9	IEEE TMI smooth FIR article. Page 7	167
E.10	IEEE TMI smooth FIR article. Page 8	168
E.11	IEEE TMI smooth FIR article. Page 9	169
E.12	IEEE TMI smooth FIR article. Page 10	170
E.13	IEEE TMI smooth FIR article. Page 11	171
E.14	IEEE TMI smooth FIR article. Page 12	172
E.15	IEEE TMI smooth FIR article. Page 13	173
E.16	IEEE TMI smooth FIR article. Page 14	174
E.17	NeuroImage clustering article. Page 1	175
E.18	NeuroImage clustering article. Page 2	176
E.19	NeuroImage clustering article. Page 3	177
E.20	NeuroImage clustering article. Page 4	178
E.21	NeuroImage clustering article. Page 5	179
E.22	NeuroImage clustering article. Page 6	180
E.23	NeuroImage clustering article. Page 7	181
E.24	NeuroImage clustering article. Page 8	182
E.25	NeuroImage clustering article. Page 9	183
E.26	NeuroImage clustering article. Page 10	184
E.27	NeuroImage clustering article. Page 11	185
E.28	NeuroImage clustering article. Page 12	186
E.29	NeuroImage clustering article. Page 13	187
E.30	HBM'98 abstract: Canonical ridge analysis	188
E.31	NeuroImage PCA article. Page 1	189
E.32	NeuroImage PCA article. Page 2	190
E.33	NeuroImage PCA article. Page 3	191
E.34	NeuroImage PCA article. Page 4	192
E.35	NeuroImage PCA article. Page 5	193
E.36	NeuroImage PCA article. Page 6	194
E.37	NeuroImage PCA article. Page 7	195
E.38	NeuroImage PCA article. Page 8	196
E.39	NeuroImage PCA article. Page 9	197
E.40	NeuroImage PCA article. Page 10	198
E.41	NeuroImage PCA article. Page 11	199
E.42	NeuroImage plurality article. Page 1	200
E.43	NeuroImage plurality article. Page 2	201
E.44	NeuroImage plurality article. Page 3	202
E.45	NeuroImage plurality article. Page 4	203
E.46	NeuroImage plurality article. Page 5	204
E.47	NeuroImage plurality article. Page 6	205
E.48	NeuroImage plurality article. Page 7	206
E.49	NeuroImage plurality article. Page 8	207
E.50	NeuroImage plurality article. Page 9	208
E.51	NeuroImage plurality article. Page 10	209
E.52	NeuroImage plurality article. Page 11	210
E.53	NeuroImage plurality article. Page 12	211
E.54	NeuroImage plurality article. Page 13	212
E.55	NeuroImage plurality article. Page 14	213
E.56	NeuroImage plurality article. Page 15	214
E.57	NeuroImage plurality article. Page 16	215
E.58	NeuroImage plurality article. Page 17	216
E.59	NeuroImage plurality article. Page 18	217
E.60	NeuroImage plurality article. Page 19	218
E.61	NeuroImage plurality article. Page 20	219

E.62 NeuroImage plurality article. Page 21	220
E.63 NeuroImage plurality article. Page 22	221
E.64 NPIVM'97 article. Page 1	222
E.65 NPIVM'97 article. Page 2	223
E.66 NPIVM'97 article. Page 3	224
E.67 NPIVM'97 article. Page 4	225
E.68 HBM'98 abstract: VRML in Neuroinformatics	226
E.69 VDE2000 article. Page 1	227
E.70 VDE2000 article. Page 2	228
E.71 VDE2000 article. Page 3	229
E.72 VDE2000 article. Page 4	230
E.73 VDE2000 article. Page 5	231
E.74 VDE2000 article. Page 6	232
E.75 Unpublished article: BrainMap modeling. Page 1	233
E.76 Unpublished article: BrainMap modeling. Page 2	234
E.77 Unpublished article: BrainMap modeling. Page 3	235
E.78 Unpublished article: BrainMap modeling. Page 4	236
E.79 Unpublished article: BrainMap modeling. Page 5	237
E.80 Unpublished article: BrainMap modeling. Page 6	238
E.81 Unpublished article: BrainMap modeling. Page 7	239
E.82 Unpublished abstract: BrainMap modeling	240
E.83 HBM'2001 abstract: BrainMap outliers. Page 1	241
E.84 HBM'2001 abstract: BrainMap outliers. Page 2	242
E.85 Human Brain Mapping article. Page 1	243
E.86 Human Brain Mapping article. Page 2	244
E.87 Human Brain Mapping article. Page 3	245
E.88 Human Brain Mapping article. Page 4	246
E.89 Human Brain Mapping article. Page 5	247
E.90 Human Brain Mapping article. Page 6	248
E.91 Human Brain Mapping article. Page 7	249
E.92 Human Brain Mapping article. Page 8	250
E.93 Human Brain Mapping article. Page 9	251
E.94 Human Brain Mapping article. Page 10	252
E.95 Human Brain Mapping article. Page 11	253
E.96 Human Brain Mapping article. Page 12	254
E.97 Human Brain Mapping article. Page 13	255
E.98 Human Brain Mapping article. Page 14	256
E.99 Human Brain Mapping article. Page 15	257
E.100 Human Brain Mapping article. Page 16	258
E.101 Human Brain Mapping article. Page 17	259
E.102 Human Brain Mapping article. Page 18	260
E.103 Human Brain Mapping article. Page 19	261
E.104 Human Brain Mapping article. Page 20	262
E.105 Human Brain Mapping article. Page 21	263
E.106 HBM'2001 abstract: Author cocitation. Page 1	264
E.107 HBM'2001 abstract: Author cocitation. Page 2	265
E.108 HBM'99 abstract: Lyngby	266

List of Tables

2.1	BrainMap™ “Behavioral effects”	21
2.2	Regional hemodynamics changes	25
2.3	Brain mapping techniques	27
2.4	PET beta-sources and tracers	28
3.1	Operations on a model	34
3.2	Optimization techniques	36
3.3	Model order determination and generalization criterias	40
3.4	Simultaneous inference	41
3.5	Preprocessing steps	42
3.6	Coregistration algorithms	43
3.7	Spatial normalization algorithms and software.	44
3.8	Templates: Some of the standard brains used to atlas warping	44
3.9	Segmentation	45
3.10	Identification of artifacts in fMRI.	46
3.11	Modeling and removal of artifacts	46
3.12	Normalization methods	48
3.13	Clustering in functional neuroimaging	52
3.14	Specialization of Jöreskog’s model	64
3.15	Hemodynamic response models: Linear and nonlinear	65
3.16	Examples of nonlinear models	69
3.17	Partial least squares algorithms	73
3.18	Canonical analysis in functional neuroimaging	80
4.1	Visualization tools in brain mapping	90
5.1	Web-based neuroinformatics tools	94
5.2	Meta-analyses and reviews	98
5.3	Techniques for machine text analysis.	99
5.4	Weighting functions for frequency-based vectors	102
5.5	Author citations	106
5.6	Variation in journal naming	115
5.7	Number of articles in neuroscience journals	116
5.8	Finding related authors	119
5.9	Authoritative authors on the term “fmri”	119
5.10	BrainMap™ outliers	123
5.11	Count of the different number of phrases the word "lobe" appear in.	124
A.1	Brodmann numbers	143
A.2	BrainMap™ “Behavioral effects”	144

Chapter 1

Introduction

1.1 What is functional neuroimaging?

Functional neuroimaging aims to understand the link between anatomical brain locations and psychological functions and especially deals with establishing pictures of the brain. Functional neuroimaging techniques include positron emission tomography (PET) or functional magnetic resonance imaging (fMRI).

1.2 Why functional neuroimaging?

- **Scientific exploration and understanding of the brain.** Functional neuroimaging is often regarded as fundamental science that has no direct goals and concerned with “truth” more than “usefulness”.
- **Direct clinical application**
 - **Neurosurgery.** One area often mentioned as an application of functional neuroimaging is preoperative and intraoperative brain mapping for surgery guidance. In preoperative brain mapping important areas of the brain are identified prior to a neurosurgical operation, e.g., it is important to identify in which hemisphere the language resides and the usual methods has been the so-called “Wada-test” (Wada and Rasmussen 1960) that is very invasive for the patient. Functional neuroimaging can be less invasive and more accurate¹
 - **Diagnostic.** Functional neuroimaging can provide objective measurements of mental activities that previously only have been available to the medical doctor/psychologist through the “subjective” account of the patient, which is useful, e.g., in connection with examination hysterical patients (Davis, Giannoylis, Downar, Kwan, Mikulis, Crawley, Nicholson, and Mailis 2001). Infants that cannot communicate their perception abilities can be examined by functional neuroimaging with a possible predictive value for later perception performance (Born, Miranda, Rostrup, Toft, Peitersen, Larsson, and Lou 2000b).
- **Artificial intelligence.** Understanding how the brain works can help artificial intelligence in developing more advanced algorithms that have practical importance in many areas of the engineering sciences. Models of biological neuronal network have inspired the development of artificial neural networks that have been applied in numerous fields. Deeper understanding of the human visual system might help computer vision, understanding how the brain handles language might help develop computer programs with better natural language capabilities and understanding of the binding problem might help database development.
- **Methodology development.** If a question is difficult enough the task of answering it could foster development of new methods, that in turn can be of use in other technical or scientific areas. The questions in functional neuroimaging are often quite difficult and the field has a large methodological subfield. Some of the methods developed can be of use in other areas: An example is the development of tests in random fields, with application in astrophysics (Worsley 1995).

¹Posner and Raichle (1995) describe that Pardo and Fox (1993) found that the assessment of the dominant hemisphere for language was performed more accurately with PET than with the Wada test, as there was discrepancy between the PET and Wada test in one of the nine subjects in the study, — and the PET was correct in that single case. Assessment of language lateralization for preoperative evaluation has also been performed with fMRI, see e.g. Desmond et al. (1995), Xiong et al. (1998), Lin et al. (1997), Brockway (2000), Hund-Georgiadis et al. (2001), J. E. Adcock and Matthews (2001). Kennan and Constable (2001) describes the assessment using NIRS

- **Brain-computer interfaces.**

- **Brain to computer interface (BCI).** This type of interface enables the brain (e.g. in patient with total motor paralysis) to control computers or machines. A system based on a neurotrophic electrode implanted “into the outer layers of the human neocortex” made a patient able to control a cursor on a computer monitor (Kennedy et al. 2000). fMRI was used to identify a suitable area of placement (motor cortex). Other techniques use surface electrodes with, e.g., P300-detection or “slow cortical potentials” (SCP) (Kübler et al. 1999). The output is still limited to very few parameters: x- and y-coordinates, and “enter”, and systems with higher “bandwidth” has probably a long way to go yet.
- **Computer to brain interface.** As a brain-computer interface can help patients with motor disabilities, a computer-to-brain interface can help patients with perception disabilities. An example (for peripheral nerves) is cochlear implants that stimulate the nerve fibers in the inner hear of people with profound or total hearing loss (Brown 1999, page 796–797). Commercial systems exist with 22 electrodes that receive their signal from a small digital signal processing computer with an attached microphone. Dobelle (2000) describes a visual prosthesis system — the “Dobelle Eye” — which features a sub-miniature television, a processing computer and an array of 68 electrodes implanted on the surface of the visual cortex.

1.3 Computers, mathematics and statistics in functional neuroimaging

Computers, mathematics and statistics play an important role in functional neuroimaging. At the very basic level tomographic brain scanners such as PET, CT and MRI rely on mathematical reconstruction for the production of the brain image, and the data sets are so large that this reconstruction is only feasible with the help of computers. The ways in which computer engineers can contribute to functional brain mapping can be grouped in:

- Development of new mathematical and statistical methods to process and analyze functional neuroimaging data and development of useful tools that the neuroscientist can immediately handle.
- Development of computer visualization techniques for visualization of the functional neuroimaging data.
- Development of database tools for searching, comparison and evaluation across experiments.

The term *neuroinformatics* has been used to denote the field that are concerned with these issues.

The body of research in functional neuroimaging and the related fields in cognitive and neuroscience are becoming so large (e.g., the exponential rise in citations to the Talairach atlas, Fox 1997) that it is difficult for a human to navigate the data without the support of computers, — in the words of Bush (1945)

The investigator is staggered by the findings and conclusions of thousands of other workers – conclusions which he cannot find time to grasp, much less to remember, as they appear.

This writing inspired the development of the world wide web of the Internet (Berners-Lee, Cailliau, Groff, and Pollermann 1992; Berners-Lee, Cailliau, Luotonen, Nielsen, and Secret 1994), that in connection with functional neuroimaging is used for search and retrieval of textural information such as scientific literature, as well as distribution of software and neuroscientific data. Perhaps the most important issue with databases and other information interfaces is knowledge access efficiency (Pitkow 1997, page 7):

One goal of information interfaces is to maximize user interaction by increasing the amount of accessible knowledge in shorter periods of time.

Databases containing neuroscientific data are still scarce and it is not clear what information is relevant to collect. Indeed one of the tasks of neuroinformatics is the *definition* of useful information. The utility of the databases is not only determined by their ability of organizing data, but also in enabling collaborations between researcher of different skills, i.e., experimental/observational and theoretical/analytical, e.g., Tycho Brahe’s astronomical database (Brahe 1602; Kepler 1627) formed the basis for the development of Johannes Kepler’s model (Kepler 1609; Kepler 1619).

1.4 Contribution

On the descriptive level the main contributions of this thesis are:

- Overview of preprocessing for functional neuroimages.
- Overview of analysis of functional neuroimages. This should be useful as a description of the algorithms in the Lyngby Matlab toolbox (Hansen et al. 1999b), (Hansen et al. 2000b), (Hansen et al. 2001b), and can act as a companion to the manual (Toft, Nielsen, Liptrot, and Hansen 2001).

- Overview of visualization of functional neuroimages.
- Introduction of neuroinformatics.

Apart from the direct application as a reference the two first overviews can also form the basis for a discussion on what information should be collected in neuroinformatics databases for functional neuroimaging.

On the level of analysis the main contributions in this thesis are:

- General canonical analysis model. Identification of a type of partial least squares as a canonical ridge analysis, with the possibility of interpolating between a canonical correlation analysis and a partial least squares solution.
- VRML visualization in connection with functional neuroimaging.
- Analysis of functional neuroimaging locations with flexible probability density modeling for outlier detection. Specifically this means finding outliers in the BrainMapTM database via kernel density estimation.
- Author cocitation and journal cocitation analysis, specifically applied in functional neuroimaging with the analysis of a single journal “NeuroImage”. In this connection term and link based analysis is combined to what is here called *context sensitive author ranking*. Author cocitation analysis was first described by White (1981), but to my knowledge it has not been based on data directly obtained from the publisher’s website and link and text based analysis has not been combined in connection with scientific documents.

1.5 Outline

Chapter 2 describes the objects under study: The brain and its associated mental processes, how they are related and how it is possible to measure the functional activity.

Chapter 3 describes the computerized and mathematical analysis of functional neuroimages, beginning with general principles with no particular reference to analysis of functional brain mapping (models, estimation and testing), then describing *preprocessing* of functional neuroimages and finally describing of some of the individual mathematical models encountered in functional neuroimaging analysis.

Chapter 4 describes the scientific visualization of 2D and 3D functional neuroimages and information visualization in connection with functional neuroimaging research.

Chapter 5 introduces neuroinformatics and focus on text analysis and (meta-)analysis of activation foci from functional neuroimaging experiments.

Abbreviations are expanded in section A.3 page 140, and some of the expanded terms can be looked up in the word list in section A.2 page 134. The abbreviations and words are divided between neuroanatomical names and other general names. A few derivations appear in section B, and acknowledgment is on page 155.

The bibliography starting on page 267 is with URLs, Pubmed identifiers (PMID) and with links to ResearchIndex (see section 5.2.5). The URLs might be invalid but most of these invalid URLs should be locatable with the suggestion provided by Lawrence et al. (2001).

Author index and ordinary index begin on pages 319 and 337, respectively.

Chapter 2

Brain, mind and measurements

The following sections introduce the information structures used in functional neuroimaging for organizing the mind/behavior and the brain. The coupling between the mind/behavior and brain is further described, and finally techniques for in vivo measurement of the brain are discussed.

2.1 Behavioral and cognitive components

A *cognitive component* is a separate entity of the mind. The cognitive component might in the first instance just be a heuristic classification by the researcher, — in the second instance we might hope that the cognitive component constitutes a “true” class. The cognitive components can be grouped in major cognitive components, e.g., perception, motion, cognition, and emotion. These may consists of subgroups that in turn consists of “mind atoms” or *core processes*. Brain mapping (and cognitive psychology) aims at identifying what the appropriate cognitive components are, their relations and how they are related to brain areas.

The term *behavioral components* can be used to denote a broader class of mental processes, states, stimuli and responses, e.g., neurological diseases and pharmacological stimuli, which would incorporate the variables that are of interest to collect in a neuroscientific database. Some of the early attempts in classification of behavioral components were made by Galen (129–199 AD) with the four temperaments (sanguine, phlegmatic, melancholic and choleric) and Franz Joseph Gall (1758–1828) with the 27 “faculties” (Gade 1997). A modern system of cognitive/behavioral components is the hierarchy in the BrainMap™ database (see section 5.2.1) shown in figure 2.1. The organization of the cognitive

Type	Subtype	Example
Perception	Audition	Noise
	Gustation	Salt
	Olfaction	
	Somethesis	Pain
	Vision	Motion
Motion	Execution	Hand flexion
	Music	
	Preparation	Articulatory coding
	Speech	Word repetition
Cognition	Attention	Divided
	Language	Phonology
	Mathematics	
	Memory	Primed words
Emotion		Anxiety
Disease		Depression
Drug		Apomorphine

Table 2.1: BrainMap™ “Behavioral effects” — a hierarchy of behavioral and cognitive components. More detail is available in table A.2.

components is not necessarily a tree-structure, e.g., “the mathematical cognitive component” might consists of a language circuit associated part (in the left frontal lobe) and a visuo-spatial associated part (Dehaene, Spelke, Pinel, Stanescu, and Tsivkin 1999; Simon 1999; Spelke and Dehaene 1999). Thus mathematical processing can be see as part of a language cognitive component and of a visuospatial cognitive component, see figure 2.1.

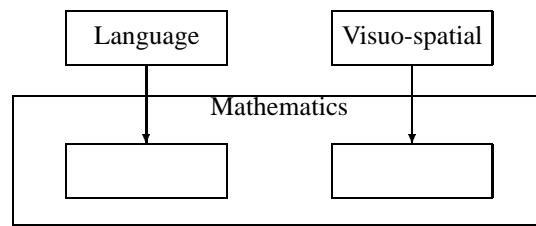


Figure 2.1: Segregation of mathematics cognitive components — “numerical thinking” as suggested by Spelke and Dehaene (1999). The visuo-spatial non-linguistic numerical processing are believed to occur in the inferior parietal cortices bilaterally.

In BrainMap™ “mathematics” presently has its own group under “cognition”. Furthermore, obsessive-compulsive disorder (as in e.g. Rauch, Jenike, Alpert, Breiter, Savage, and Fischman 1994) is categorized under “emotion” while it might also be appropriate to group it under “disease”. These examples suggest that a simple tree structure for organizing cognitive components is too restrictive. Perhaps a direct acyclic graph is sufficiently general to described the relationship between the components.

When a *task* is performed or a *stimulus* is presented (collectively, any elicitation of mental processes during an experiment whether internal or external generated is often referred to as the *paradigm*) one or more cognitive components are involved. These may be engaged in parallel or serial, — as in the word production model of Indefrey and Levelt (2000) shown in figure 2.2, e.g., the task of silent picture naming may involve all but the last two core processes.

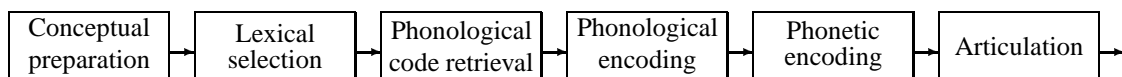


Figure 2.2: Core processes cognitive components involved in word production as proposed by Indefrey and Levelt (2000, figure 59.2).

Though functional networks can be revealed by measuring a subject in *one* (resting) state (Lowe et al. 2000), a typical functional neuroimaging experiment will try to elicit a single cognitive component by exposing subjects to two task: one where the cognitive component appear (the “activation”) and another where it does not appear (the “baseline” or “control condition”). The brain activations are measured in both states and the two measurements are (in some way) subtracted from each other — the so-called *cognitive subtraction paradigm* (Friston et al. 1996c; Law 1996, section 1.6.2). It is often difficult to find two tasks that will mask out the appropriate cognitive component. Often simple so-called “rest” states are used as the baseline where the subject is supposed to do “nothing”, — usually either with closed eyes or fixation on a target. However, when the subject is “doing nothing” s/he might be engaged in “semantic knowledge retrieval, representation in awareness, and directed manipulation of represented knowledge for organization, problem-solving, and planning” and the performance of the task can be interpreted as a interruption of such “conceptual” processes (Binder et al. 1999). Along a similar line Gusnard, Akbudak, Shulman, and Raichle (2001) find that medial prefrontal cortex is engaged in “self-referential mental activity”, see also (McGuire, Paulesu, Frackowiak, and Frith 1996). Thus it can be an advantage to engage the subject in an extra task that requires “cognitive load”. It is, on the other hand, convenient for meta-analysis studies that the same baseline is used across studies.

If two (or more) cognitive components are investigated simultaneously one can investigate the *interaction* effect with a 2×2 factorial design. If it is not possible to factor out both components *cognitive conjunction* can be used (Price and Friston 1997).

It is interesting to note that a task can be regarded as a mix of core processes. Furthermore, if a cognitive component should constitute a core process it would be useful to regard it as independent from other core processes. Under these assumption the identification of core processes can be see as *independent component analysis*, see section 3.11.

2.2 Specialization of the Brain

The mental processes engage the neurons of the brain. The neurons are special cells with neurites (dendrites or axons) which are long projection from the cell body that connect to other neurons through junction points called *synapses*. Areas with high density of neuron bodies are referred to as gray matter (GM) and areas with high density of connections are referred to as white matter (WM). These connections are usually aggregated in bundles. Other areas of the brain contain cerebrospinal fluid (CSF) and vessels.

There have been established standardized hierarchies for the macroscopic anatomy of the brain: National Library of Medicine has as a service: The “MeSH Browser” which contains some of the major parts of the brain organized in a hierarchy (<http://www.ncbi.nlm.nih.gov:80/entrez/meshbrowser.cgi>). Another effort is NeuroNames (Bowden and Martin 1995) which main components of the hierarchy are displayed in the list in figure 2.3. The hierarchy can further be divided, e.g., according to the sulci and gyri of the cerebral cortex. The precise appearance and location of these sulci and gyri varies among individuals (Ono, Kubik, and Abernathey 1990).

```

Forebrain
  Telencephalon
    Cerebral cortex
      Frontal lobe
      Parietal lobe
      Insula
      Temporal lobe
      Occipital lobe
      Cingulate gyrus
      Parahippocampal gyrus
      Archicortex
        Supracallosal gyrus
        hippocampal formation
    Cerebral white matter
    Lateral ventricle
    Basal ganglia
      Striatum
      Globus pallidus
      Amygdala
    Septum
    Fornix
  Diencephalon
    Epithalamus
    Thalamus
    Hypothalamus
    Subthalamus
    Third ventricle
Midbrain
  Tectum
  Cerebral peduncle
    Mindbrain tegmentum
    Substantia nigra
Hindbrain
  Metencephalon
    Pons
    Cerebellum
  Medulla oblongata

```

Figure 2.3: Main components of NeuroNames brain Hierarchy (Bowden and Martin 1995). <http://rprcsgi.rprc.washington.edu/neuronames/index1.html>

Correlation between cognitive variables and the static appearance of brain structure has been found, e.g., Maguire et al. (2000) showed that the posterior hippocampi were larger (and the anterior hippocampal region smaller) in taxis drivers compared with a group of controls, suggesting the anterior hippocampus being involved in navigation.

There exists several brain atlases that renders the brain in a coordinate space, so-called stereotactic space. The atlas of Talairach and Tournoux (1988) has been particular used and many functional neuroimaging studies report results in the coordinate system set up in this atlas.

2.2.1 Microscopic structure

There is a number of criteria for classifying brain areas on the microscopic level. This can be done from cyto-, myelo-, or receptor architectonic criteria, see e.g., (Zilles and Palomero-Gallagher 2001).

Cytoarchitectonical maps can be made of the brain by classification of the appearance of neurons, their network and density. It is usually performed by examining the cells in a microscope after staining of the cellular components, — e.g., the cell bodies as in the *Nissl method* (Heimer 1994, section 7). The Brodmann classification is a widely used cytoarchitectonic classification system for the cerebral cortex (Brodmann 1909). It delineates what has now been termed *Brodmann areas* (BA) and assign *Brodmann numbers* from 1 to 47 to each region, see table A.1 on page 143. Another cytoarchitectonic classification less widely used is that of von Economo (1929). Other classifications of the brain on the anatomical level can be made by receptor-based maps; these are correlating well with the cytoarchitectonics maps (Geyer, Schleicher, and Zilles 1997). On the other hand will the cytoarchitectonic classification usually not be related to sulci structure (Roland et al. 1997; Rorden and Brett 2000; Amunts et al. 1999; Amunts et al. 2000; Morosan et al. 2001), and

large intersubject variability exists, e.g., “the volumes of area 44 differed across subjects by up to a factor of 10” (Amunts et al. 1999).

The Talairach atlas (Talairach and Tournoux 1988) marks the Brodmann areas, and the Talairach Daemon (Lancaster, Summerlin, Rainey, Freitas, and Fox 1997c) is able to translate between stereotactic coordinates and Brodmann numbers. The cytoarchitectonic observations are based on non-living material as no technique has been developed for imaging the cytoarchitectonic structure in vivo.

2.2.2 Functional specialization

Even though the mental processes can be split into separate cognitive components it does not mean that the cognitive components are specialized in specific brain regions. The first evidences that the brain is specialized — at least to some extent — were the studies of patients with aphasia by Pierre Paul Broca and Wernicke (Broca 1960). Later, functional specialization was also found between the left and right hemisphere with split brain patients, see e.g., Shepherd (1994, p. 678–679).

Brain mapping rests on the paradigm that the brain *is* specialized — at least to some degree. How much is still a controversy: In one end the most radical would argue that the brain is segregated into discrete regions and each region performs one unique cognitive task. Any fuzziness that is seen in brain mapping is due to limitations in the measurements, e.g., low image resolution in brain scanners, the filtering from the neurovascular coupling or subject variations. This view is called “locationistic”. At the other end is the view that all areas of the brain are participating in all cognitive tasks just with differing degrees. This view is sometimes called “connectionistic” and other phrases used in this domain is “parallel distributed processing” and “integrated networks”.

The two views differ in what they believe the result of a functional neuroimage analysis should be: The “discrete segregation” view holds that the result should be a labeled volume, with each unique label referring to a cognitive component. The center of mass of a connected region with the same label can be extracted and put into a table that is publishable. The “distributed” view holds that the result should be a vector for each cognitive component with values for each voxel indicating its “degree of involvement” in the cognitive task.

There is not necessarily a one-to-one mapping between a cognitive component and a brain area: One cognitive component might be “implemented” in two or more different brain areas (this is referred to as degeneracy, e.g. by Price and Friston 1999), and one brain area might process two or more cognitive processes. An example of the latter is the syntax processing in the Broca’s area that both processes linguistic syntax processing as well as musical syntax processing (Maess, Koelsch, Gunter, and Friderici 2001), though linguistic syntax processing is normally confined to the left hemisphere, while music processing is found in the right hemisphere.

Do all cognitive components have spatial specialization? Probably not, but even (general) intelligence has been correlated with a specific brain region (the lateral frontal cortex) (Duncan et al. 2000).

2.3 Coupling

Mental activity manifests itself as electrochemical activity in the neurons of the brain: When a neuron “fires” it changes its electric potential over a period of milliseconds and the spike travels along the axon, — the so-called action potential. When the signal is transferred from the axon via the synapse a postsynaptic potential is generated on the receiving neuron. The postsynaptic potentials can be excitatory or inhibitory. If about 100 excitatory synapses, on average, get activated on a single cell within in a short time interval they trigger a new action potential (Longstaff 2000).

The electrochemical activity in turn requires energy causing a number of physiological changes in the brain. Angelo Mosso in 1878 and Roy and Sherrington (1890) were the first to discover a relationship between mental activity and a physiological response. Later Kety and Schmidt (1945) were able to measure the global cerebral blood flow, and Lassen, Ingvar, and Skinhøj (1978) created the first 2-dimensional activation images.

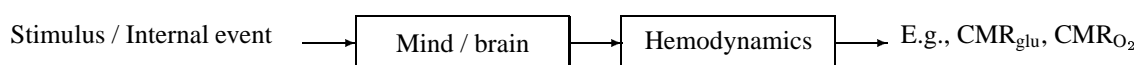


Figure 2.4: Coupling

As shown in figure 2.4 the coupling can be thought of as consisting of two stages which will be described below: a stimulus-neural coupling and a hemodynamic coupling.

Component	Increase	Comment
rCMRglu	20-40%	Mostly oxydative glucolysis
rCMRO2	5-25%	
rCBF	20-70%	Increases in velocity rather than capillary recruitment
rCBV	5-30%	Mostly in the venous vessels.

Table 2.2: Regional hemodynamic changes with activation. From Jezzard (1999). Hoge et al. (1999) get “CBF and CMRO2 increases of $48 \pm 5\%$ and $25 \pm 4\%$, respectively”.

2.3.1 Stimulus-neuronal coupling

How does the neuronal activation depend on cognitive processing? If an stimulus is imposed on the brain does it then respond linearly?

For stimuli that activate primary sensory area there might be a one-to-one mapping in the timing between the stimulus and the neuronal response. However, for more complicated tasks the neuronal response is not necessarily one-to-one associated in the timing with external events, e.g., Konishi et al. (2001) finds activation associated with the transitions in a block design.

Another complexity might appear in connection with learning and memory where the neuronal response are on a longer time scale than the stimulus, e.g., memory consolidation might happen during sleep (Graves, Pack, and Abel 2001) many hours after the stimulus occurred.

2.3.2 Hemodynamics: Coupling between neuronal activity, blood flow and metabolism

The electrochemical activity of the neurons requires energy mainly in connection with Na^+ and K^+ -ATPase activity, and almost all neuronal energy derives from oxydative glucose metabolism (Jezzard 1999). A smaller part derives from non-oxydative (anaerobic) metabolism (Prichard et al. 1991). The glucose uptake is not completely correlated with the neural activity since there is a higher uptake than required by metabolism (Fox and Raichle 1986; Fox, Raichle, Mintun, and Dence 1988). There is evidence that the glucose is stored and used later during non-activation (Madsen et al. 1999; Madsen 2000), see also (Barinaga 1997) for an introduction to these matters.

Apart from an increased glucose (CMRglu) and oxygen (CMRO2) metabolism neural activation results in increased blood flow (CBF), blood volume (CBV) and blood oxygenation. The CBF usually increases more than is “needed”, — the phenomenon termed “uncoupling” or “luxury perfusion” (Fox and Raichle 1986; Buxton and Frank 1997). The amount of oxyhemoglobin (HbO_2) and deoxyhemoglobin (Hbr) is a function of the CBF, CMRO2 and CBV, see Table 2.2. Thus the BOLD response of fMRI (see section 2.4.5) which is dependent on HbO_2 and Hbr is a function of several variables that each can have different time and spatial behavior. These variables are in turn dependent on other variables of the brain. Bandettini et al. (1993) list the activity variables that are likely to influence the coupling between stimulus and BOLD response: blood pressure, hematocrit (Hct), blood volume, blood pO_2 (oxygen partial pressure), perfusion rate, vascular tone (amount of vasodilatory capacity), neuronal metabolic rate, vasodilation (enlargement of blood vessel), blood oxygenation, blood perfusion.

Aguirre, Zarahn, and D’Esposito (1998b) and Glover (1999) find that the fMRI response varies across subjects so that an individual model should be used for each subject. Intrasubject intersession non-stationarities are seen by McGonigle et al. (2000) and Miki et al. (2000). This is supported by “voxel counting” across trials and subjects by Cohen and DuBois (1999). Rostrup et al. (2000) found a “considerable difference” between the response in WM and GM.

The form of the hemodynamic response function

The hemodynamic response is filtered in space and time: An infinitely short and point-like neuronal activation will elicit hemodynamic response that lasts some seconds and is dispersed some millimeters. A temporally linear and stationary model for the hemodynamic response is implemented in SPM and the impulse response to that filter is shown in figure 2.5. This model was originally found in a study by (Glover 1999). A linear stationary model cannot implement different shapes for the onset and the cessation from a block stimulus — they should have the same (mirrored) curve. Bandettini et al. (1993) mention that the delay in signal change is 5-8 seconds from stimulus onset to 90% maximum and 5 to 9 seconds from stimulus cessation to 10% above baseline. If there is a discrepancy between the onset and cessation then a linear and stationary model is only an approximation.

The main components in the impulse response function (IRF) of the hemodynamic response (as seen in BOLD fMRI) are a positive peak around 5 seconds and a post-activation undershot (post peak dip) approximately at 15 seconds, see figure 2.5. The IRF obtained empirically by Zarahn, Aguirre, and D’Esposito (1997, figure 4) has the maximum at 6 seconds and the post-activation undershot minimum at 13 seconds. These values approximately corresponds to the filter found by Goutte, Nielsen, and Hansen (2000, figure 6a).

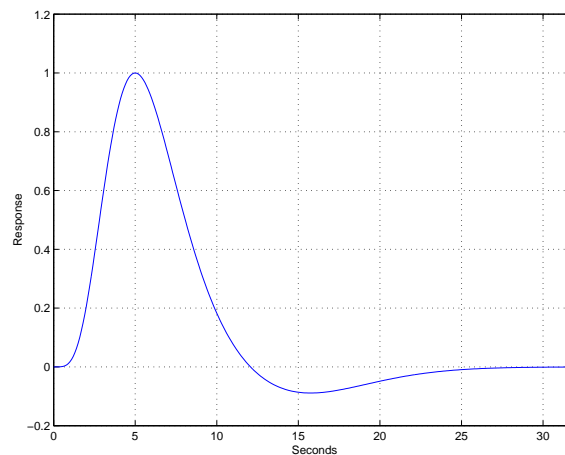


Figure 2.5: The “Glover” impulse response function for a simple linear and stationary model of the hemodynamic response function (Glover 1999; Friston 1999a). The minimum is at 15.7 seconds and the maximum at 5 seconds.

More components have been attributed to the hemodynamic response function. One of the most interesting is the so-called “initial dip” (early dip or fast response) which is a negative response that should occur in the few first seconds of the response with a peak after approximately 2 seconds (Yacoub and Hu 2001). It is interpreted as a quick local oxygen metabolism that is not compensated for by an increased rCBF which results in an increase in deoxyhemoglobin, detected by the fMRI scanner. The initial dip should be more localized than the rest of the hemodynamic response as the deoxygenation is more local than rCBF. Thus the initial dip promise higher spatial and temporal resolution than the ordinary response. Unfortunately the effect is not very large, and it is not seen in all experiments. Some of the positive reports are: imaging spectroscopy (Malonek and Grinvald 1996), fMRI (Ernst and Hennig 1994), fMRI at 4T (Menon, Ogawa, Hu, Strupp, Anderson, and Ugurbil 1995), fMRI at 1.5T (Yacoub and Hu 1999), dependence on TE (Yacoub, Le, Ugurbil, and Hu 1999), initial dip in motor and visual areas (Yacoub and Hu 2001), initial dip with oxygen dependent phosphorescence quenching optical imaging (Vanzetta and Grinvald 1999). For a discussion of the divergent results on detection of the initial dip see Buxton (2001) and Vanzetta and Grinvald (2001). A physiological model — the so-called “balloon model” can account both for the postactivation undershoot and the initial dip (Buxton, Wong, and Frank 1998) (see (Glover 1999)).

Apart from the short term response (scale of seconds) there are also long term effects on the scale of minutes. Krüger, Kleinschmidt, and Frahm (1996) report an initial overshoot, signal decrease extending over 4-5 min, post-activation undershoot (of response) mirroring of initial overshoot. Jones (1999) reports that the post-activation undershoot can last up to 1 minute.

2.3.3 Deactivation

In some functional neuroimaging studies *deactivation* is seen. This is not directly related to increased activity in inhibitory synapses since they also require energy for their activity. However, inhibition will presumably cause a decrease in the total number of firings.

A negative BOLD response can be found in sedated (pentobarbital, chlor-hydrate) and anaesthetised (helothane/nitrous oxide) children (Born et al. 1996; Joeri et al. 1996). Born et al. (2000b) also found a negative BOLD response among some young sleeping or sedated (chloral hydrate) children (while Hykin et al. (1999) demonstrated *positive* BOLD-signal in the fetus brain, when applying a auditory stimulus). The negative BOLD response is not restricted to sedated children but can be found in sleeping adults (Born et al. 2000a). The negative response is also seen in rCBF PET (Born et al. 2001).

Artifactual deactivations can be seen if confounds are modeled incorrectly, see section 3.5.8, and if subjects engage in paradigm-unrelated mental processes during “rest” scans, see section 2.1.

2.3.4 Coupling nonlinearity

The hemodynamic response on the scale of seconds is found to be approximately linear if the stimuli is sufficiently long (Boynton, Engel, Glover, and Heeger 1996). However, for shorter stimuli nonlinearities are found: (Vazquez and Noll 1998), auditory stimuli shorter than 6 seconds (Robson, Dorosz, and Gore 1998), speech syllables (Binder, Rao, Hammeke, Frost, Bandettini, and Hyde 1994), rapidly presented nouns (Friston, Josephs, Rees, and Turner 1998b), metronome paced finger tapping (Glover 1999) and visual checkerboard stimuli shorter than 4 seconds and with varying on and off periods (Birn and Bandettini 2001).

Abbrv.	Exp.	Name
CT/CAT	0	Computerized (axial) tomography
MRI/fMRI/pMRI/MRS/MRSI	18	Magnetic resonance imaging
PET	619	Positron emission tomography
SPECT	1	Single photon emission computed tomography
fNIR/fNIRS/NIRS	0	Functional near-infrared spectroscopy
TCD	0	Transcranial Doppler
EEG/ERP	8	Electroencephalography / Event-related potentials
MEG	0	Magnetoencephalography
EIT	0	Electrical Impedance Tomography
ECoG/SEEG	0	Electrocorticography / Stereoelectroencephalography
LFP/FP	0	(Local) field potential
OI	0	Optical imaging with voltage sensitive dyes
OIS	0	Optical intrinsic signal imaging
LD	0	Laser-Doppler
	0	Single-cell/Multiple-cell electrophysiology
ESM	0	Electrocortical stimulation mapping / Electrical corticostimulation
TMS	0	Transcranial magnetic stimulation

Table 2.3: Brain mapping techniques (modality). The “Exp.”-column shows the number of experiments in the BrainMap™ database (2000 May) being marked as performed with the modality, — some of these experiments combining two modalities.

Furthermore when several cognitive components are stimulated simultaneously the effect might not be additive.

2.4 Functional neuroimaging and other brain measurement techniques

Almost all experiments recorded in the BrainMap™ database use either PET or fMRI, see second column in table 2.3. A few use EEG (ERP) in combination with PET and (Allison et al. 1994) is the only study where Talairach coordinates are given based on EEG (ERP) measurements. Walter et al. (1992) uses MEG, PET and MRI (but is presently marked “PET-MRI”).

Apart from the techniques described below there are among others *transcranial Doppler (ultrasound sonography)* (TCD) which can measure on the blood flow velocities in the cerebral arteries, and *laser-Doppler (flowmetry/velocimetry)* (LDF/LDV) which measure microcirculatory blood flow. Laser Doppler can also construct images, — so-called *laser Doppler perfusion images*.

If different modalities are combined it is possible to obtain good time resolution (with the electrical/magnetic methods) together with good spatial resolution (with, e.g., an fMRI scanner), e.g., structural MRI scans can be used to form more spatial precise MEG and EEG images (Dale and Sereno 1993) and for intersubject alignment, see section 3.5.6.

2.4.1 Electrophysiology

Direct measurement of the electric state of the single cell can be made by *patch recordings* or *intracellular recordings* with micropipettes (Shepherd 1994, pages 68–69). Multiple cells can be recorded simultaneously by extracellular recordings from a crowd of neurons and their relationship can be analyzed (Gerstein and Perkel 1972). The individual neurons can be distinguished from each other by the form of the action potential (the spike), e.g., through principal component analysis (Kirkland 2001) or some form of clustering (Rinberg, Davidowitz, and Tishby 1999).

As the measured spikes are point process-like it is not possible to analyze this data with the methods typically used with functional neuroimaging data. However, the spike train data can be converted to a spike density (in time) and discrete time samples can be represented in a matrix and analysed in the same way as PET and fMRI are.

Local field potential (LCD) is an intermediate step between electrophysiology and EEG.

2.4.2 EEG, MEG, EIT

Electroencephalography (EEG) and *magnetoencephalography* (MEG) measure the electric and magnetic field generated from a neuron assemble. In the case where the EEG signal is phase-locked to a stimulus and averaged across trials it is often called *event-related potentials* (ERP) or *evoked potentials*

Usually EEG is measured noninvasively on the scalp by an array of electrodes with the first measurement done by Hans Berger in 1924 (Berger 1929). However, there exist invasive intracranial variations of EEG: *Electrocorticography*

(ECoG) typically performed preoperative or intraoperative where an electrode grid (e.g., a 8×8 array) is placed directly on the cortical surface, and stereoelectroencephalography (SEEG) — also called “depth-EEG” or intra-cerebral recording — where the electrodes are placed deeper in the brain. These intracranial EEG techniques are usually performed in connection with neurosurgery on epileptic patients. Examples of this technique are available in (Towle et al. 1998), (Chkhenkeli et al. 1999), (Widman et al. 1999) and (Allison et al. 1994) — the only one presently recorded in the BrainMap™ database.

The ERP signal exhibits some characteristic excursions named with a combination of the polarity (negative (N) or positive (P)) and the approximate time of extremum in milliseconds, e.g., N200. The signals are sometimes denoted by the applied stimulus: auditory evoked potential (AEP), visual evoked potential (VEP), somatosensory evoked potential (SEP/SSEP) or proprioceptive evoked potential (PEP) (Arnfred et al. 2000). Rather than a signal of its own the ERP might (just) be a modulation of the alpha-rhythm of the EEG (Makeig, Jung, Westerfield, and Sejnowski 2001).

If EEG is measured from a set of electrodes it is possible to give an estimate of the 3-dimensional distribution of the electric activity or give an estimate of the optimal dipole position if a fixed number is assumed, — so-called source localization. One such approach is LORETA (Low-resolution electromagnetic tomography) (Pascual-Marqui et al. 1999).

MEG measures the very small magnetic field that arises from neuronal activity (Cohen 1968). The small signal (hundreds of femtoTesla) is measured with a so-called SQUID (super quantum interference device) and with multiple SQUIDS (presently up to several hundreds) a surface magnetic field image can be measured. As with EEG the magnetic field in the brain can be estimated. This is usually referred to as magnetic source imaging (MSI). Models with a few sources exist together with current density models, — so-called magnetic field tomography, see, e.g., (Ionnides, Bolton, and Clarke 1990). Some of the methods are synthetic aperture magnetometry (SAM), multiple signal classification (MUSIC, Mosher, Lewis, and Leahy 1992), Bayesian power imaging (BPI, Hasson and Swithenby 2000) and LORETA mentioned above.

There are EEG-related techniques that measure the peripheral rather than the central nervous system, e.g., electromyography (EMG). These are sometimes measured in connection with functional neuroimaging studies to monitor behavior, performance and confounding signals, e.g., control for eye movements can be monitored with EOG or with a video camera (as in Law 1996, page 53) and Richter, Andersen, Georgopoulos, and Kim (1997) monitor finger movement by electromyography.

Electrical impedance tomography (EIT, Holder 1987) applies a tiny current (e.g., at some frequency 200-80kHz) to a number of electrodes on the scalp and measures the static conductivity or the dynamics of the conductivity. Activation images can be obtained (Holder, Rao, and Hanquan 1996) and the static images can be used in connection with EEG and MEG source localization.

2.4.3 Computerized tomography

Computerized tomography (CT) — or more seldom computerized axial (or assisted) tomography (CAT) — uses X-rays together with Radon/Hough transform to obtain 2D or 3D images (Cormack 1963; Hounsfield 1973). It is the most used tomographic medical imaging technique (Kevles 1998, table 1), though usually used in connection with structural imaging and not used in any serious degree for functional neuroimaging. It is, however, possible to get functional images from a CT-scanner, e.g., rCBF can be measured by xenon-enhanced computed tomography (XeCT) (Hagen, Bartylla, and Piepgras 1999), and commercial systems exist that measure the perfusion and estimate CBF, CBV and MTT (mean transit time) (Eastwood 2000): so-called “Perfusion CT”. Non of the entries in BrainMap™ are made with CT.

2.4.4 Positron emission tomography and single photon emission computed tomography

Isotope	Half-life	Tracers	Exp.	Reference
O-15	122 sec	H ₂ ¹⁵ O (Water)	384	Jones, Chesler, and Ter-Pogossian (1976)
		C ¹⁵ O ₂	193	
		Butanol	23	
		¹⁵ O ₂	0(?)	
F-18	109 min	Fluoro-deoxyglucose (18-FDG)	11	E.g., Ashburner et al. (1996)
C-11	20 min	Fluoromethane	6	
		C11-flumazenil	0	
C-10	19 sec	¹⁰ CO ₂	0	Jensen, Nickles, and Holm (1998)
N-13	10 min		0	

Table 2.4: Examples of PET beta-sources and tracers. “Exp.” column shows the number of experiments performed with the tracer as recorded in the BrainMap™ database 2000 May.

Positron emission tomography (PET) and single photon emission computed tomography (SPECT) use radioactive isotopes (“radioisotopes”) and attach them to a molecule — a so-called tracer. This tracer is injected in or inhaled by

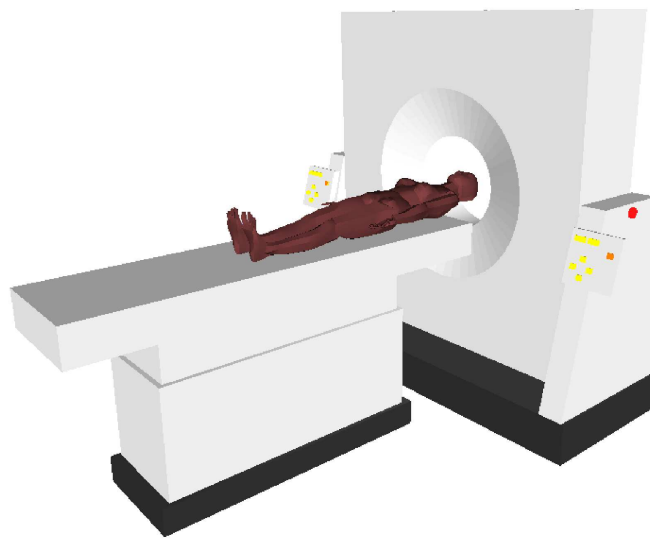


Figure 2.6: Model of a PET scanner with a subject in VRML.

the subject and distributed in the body. When the PET isotope decays it emits a positron (β^+) that after a short distance of travel annihilates with an electron. The annihilation results in a radiation of two gamma photons with an angle of approximately 180 degrees. The photons can be measured with gamma detectors in a ring around the subjects head, see figure 2.6, and the line of annihilation can be determined by a collimator or by a coincidence technique where the two gamma photons are detected with opposite detectors at the same time (Phelps 1975; Ter-Pogossian, Phelps, Hoffman, and Mullani 1975).

PET isotopes and tracers used in activation studies are shown in table 2.4. In functional neuroimaging research studies radioactive water is commonly used as it has a convenient short half-life, though a poorer spatial resolution than F-18. After injection the tracer distributes proportionally to the rCBF. SPECT scanners have a lower resolution than PET scanner and not used as much in research as PET.

A related technique is *autoradiography* where the functional brain images can be obtained on animals where radioactive tracers are injected and distributed in the brain while a stimulus is being applied. After the experiment the animal is killed and high resolution histological images can be obtained.

2.4.5 Magnetic resonance imaging (MRI)

Magnetic resonance imaging (MRI) uses a property of the atomic nucleus to make either 2D or 3D image of the brain or other anatomical parts. Atomic nucleus with either an odd number of protons or an odd number neutrons will have an observable spin (Rabi and Cohen 1933). By applying an external magnetic field B the spins can be aligned and occupy two states with different energy levels. The distribution between states can be affected externally by photons matched in frequency ν (the *resonance frequency* or *Larmor frequency*) to the energy difference $E = h\nu$, with h being the Planck constant (Purcell, Torrey, and Pound 1946; Bloch, Hansen, and Packard 1946), — the phenomenon termed *nuclear magnetic resonance*. The energy difference is dependent on the external magnetic field B and the relation to the Larmor frequency is $\nu = \gamma B$, where γ is the *gyromagnetic ratio* specific for every nucleus, e.g. 42.58MHz/T for common hydrogen H^1 . At equilibrium there are more spins aligned in the same directions as B than in the opposite direction resulting in a net magnetization from the spins.

In a simplified non-quantum mechanistic model an ensemble of spins forms a macroscopic magnetization vector. The application of the external photons (in MRI usually called the RF-pulse — radio frequency pulse) will disturb the vector and the relaxation back to its equilibrium value is associated with measurable signals, e.g., the free induction decay (FID) which decaying envelope is characterized by the $T2^*$ time constant. Other time constants are the $T1$ (longitudinal relaxation) and $T2$ (transversal relaxation) that are revealed by different RF-pulses. Apart from the decays the MR-signal can also be characterized by the amplitude of the signal which in the MRI is related to the *proton density*.

As the time constants depend on the surrounding matter (biological tissue) the non-spatial MR-signal can be used in medical application, e.g., with the potential to detect tumors (Damadian 1971). However, by varying the strength of the external magnetic field B spatially (apply the so-called *gradient field*) the Larmor frequency will vary spatially, and by controlling of the variation of the field and proper signal analysis of the relaxation signal an image can be formed (Lauterbur 1973): magnetic resonance imaging (MRI) previous termed *zeugmatography* or (spatially localized) nuclear

magnetic resonance (NMR).

By controlling the gradient field together with the application of the RF-pulses and the readout of the signals a large number of different *MRI sequences* can be made, c.f., spin echo (SE), gradient echo (GE), inversion recovery (IR), saturation recovery, FAIR, etc. Some of the parameters are: the repetition time (TR) which is the scanning period or interval between the start of each scanning, echo time (TE) which is associated with MR-sequences that use echos (inverting RF-pulses), inversion time (TI) and flip angle (FA). Apart from the modification introduced by varying these parameters the image characteristics can also be changed by intravenous contrast agent, such as gadolinium in the form of the paramagnetic Gd-DTPA, e.g., angiographic images can be constructed from pre- and post Gd-DTPA scans.

A tutorial on MRI is (Hornak 1999) and another introduction is (Leach 1988).

Functional magnetic resonance imaging

With contrast agents in the blood images can be produced that are sensitive to vascular dynamics (Villringer et al. 1988). The applied contrast agents are able to generate images of task activation (Belliveau, Kennedy Jr, McKinstry, Buchbinder, Weisskoff, Cohen, Vevea, Brady, and Rosen 1991). Hemoglobin changes its magnetic properties depending on whether it is oxygenated or deoxygenated: Deoxyhemoglobin (Hbr) is paramagnetic while oxyhemoglobin (HbO₂) is diamagnetic (Pauling and Coryell 1936). Thus hemoglobin can be used as a contrast agent — *blood oxygenation level dependent contrast (BOLD)* — for detection of oxygenation (Ogawa, Lee, Kay, and Tank 1990; Turner, Le Bihan, Moonen, Despres, and Frank 1991), and subsequently also for detection of brain activation (Ogawa, Tank, Menon, Ellermann, Kim, Merkle, and Ugurbil 1992; Kwong, Belliveau, Chesler, Goldberg, Weisskoff, Poncelet, Kennedy, Hoppel, Cohen, Turner, Cheng, Brady, and Rosen 1992).

The BOLD response is usually positive: Though oxygen is consumed and the CBV is increased the increase in CBF is larger (luxury perfusion) and the result is a decrease in Hbr with lower field inhomogeneity. However sedation/anesthesia, sleep or age can modulate it, so the response becomes negative, (Born, Rostrup, Leth, Peitersen, and Lou 1996; Born 1998; Martin, Thiel, Girard, and Marcar 2000), see also section 2.3.3.

fMRI is generally regarded as less stable than PET. This is due to instabilities of the MRI scanner, susceptibility artifacts and the complexity of the BOLD signal. Susceptibility artifacts are a special problem for experiments involving activations in the temporal region, e.g., Devlin et al. (2000b), Devlin et al. (2000a) and Veltman et al. (2000) found reduction in signal strength and fewer activations in the temporal region in fMRI compared to PET. Several compensation methods exist (Deichmann and Turner 2001). Furthermore, the fMRI signal will depend on the details of the MR-sequence, e.g., the TE (Peltier 2000).

Diffusion and Perfusion and other

Apart from BOLD MR-scanners are able to generate other types of images that are of interest in functional neuroimaging: Diffusion weighted image (DWI) are tensor images — diffusion tensor images (DTI) — (6 values for each voxels) that are able to image the laminar structure of the brain, e.g., white matter tracts can be traced (Tuch, Belliveau, and Wedeen 2000) and the thalamic nuclei can be identified (Wiegell, Tuch, Larsson, and Wedeen 2000; Wiegell 2000).

Perfusion weighted images (PWI or pMRI) are scalar time-series images (as with BOLD fMRI) and from these it is possible to obtain estimates of parameters such as CBF, CBV, so-called mean transit time (MTT) and time-to-peak. The mathematical modeling associated with PWI resembles the methods used in modeling the hemodynamic response in BOLD fMRI, e.g., (Østergaard, Weisskoff, Chesler, Gyldensted, and Rosen 1996) uses principal component regression. However, as input to the mathematical model it is rather the “input curve” — an estimate of the input of the injected contrast agent.

Furthermore, MRI can be used to measure brain temperature (Kuroda et al. 2001) and electromagnetic activity of a comparable magnitude to that of neuronal firing (Bodurka and Bandettini 2001).

2.4.6 Optical methods: Optical intrinsic signal, near-infrared spectroscopy and voltage sensitive dyes

Optical intrinsic signal

Optical intrinsic signal (OIS) (Grinvald, Lieke, Frostig, Sperling, and Wiesel 1986) — sometimes called intraoperative optical intrinsic signal (iOIS) when performed on humans during surgery — measures how white light is reflected at different wavelengths or how a bandpass filtered light is reflected on an exposed cortex. The reflected light is a function of among other variables “blood volume, blood flow, cell swelling and the oxidative state of the tissue” and the method provides one of the best spatial and temporal resolutions for functional neuroimaging: 50 μm and 50 ms (Mazziotta and Frackowiak 2000, page 18), where the oxidative (deoxygenation) signal has the highest resolution (Stetter et al. 2000).

Voltage sensitive dyes

Certain substances — so called *voltage sensitive dyes* (VSD) — change reflectance or fluorescence depending on an applied electric field. The changes can be recording by an optical imaging technique (Grinvald, Salzberg, and Cohen 1977; Grinvald, Cohen, Leshner, and Boyle 1981). Another related technique is oxygen-dependent phosphorescence, see, e.g., (Vanzetta and Grinvald 1999).

NIRS

If the frequency of the light is sufficiently low it is able to penetrate biological tissue: The non-invasive near-infrared spectroscopy (NIRS or NIROE — near infrared optical encephalography) technique uses a light source (laser diode or halogen lamp) and a spectrograph with wavelengths around 800nm (e.g., Neufang et al. (1999) analyzed the spectral range 720-940nm.) A functional activation signal can be obtained as the scatterings from deoxyhemoglobin and oxyhemoglobin are different: Wavelengths around 760nm is primarily deoxyhemoglobin and 830nm is mainly oxyhemoglobin (Chance, Chen, and Cowen 1999). The spatial resolution of NIRS is poor, but the sampling time can potentially be high, though it is usually 0.5-1 seconds (see, e.g. Kato et al. 1998). A newer study has used 6.25 Hz (Franceschini et al. 2000). NIRS in the brain was first developed by Jobsis (1977) and functional neuroimaging NIRS was first described by Kato et al. (1993)

Multisource and multidetector NIRS has also been developed. This is sometimes referred to as *optical topography* (Maki, Yamashita, Ito, Watanabe, Mayanagi, and Koizumi 2001). A similar technique that is under development is optical tomography, where “photon density is measured at various points around the medium” (George et al. 2000).

2.5 Stimulation of the brain

Apart from the stimulation through perception the brain can be stimulated directly by electric, magnetic or pharmacological means.

Electrocortical stimulation can either be applied intraoperative or during neurosurgery evaluation of epilepsy patients: An electrode grid is placed on the surface of the brain and pulses or stimulus trains can be applied with following measurements of the evoked signal (Ray, Meador, Smith, Wheless, Sittenfeld, and Clifton 1999).

Transcranial magnetic stimulation (TMS) applies a short large magnetic field on a region of the brain from a coil outside the head (Barker, Jalinous, and Freeston 1985). This causes a short virtual lesion effect. The TMS pulses can also be repeated: so-called repetitive TMS (rTMS).

Chapter 3

Analysis

3.1 Models

A *model* is a representation of a part of a world, e.g., the brain. A *mathematical model* is a mathematical representation: States and variables of the models are represented by mathematical variables or functions and inter-relation between the model states and variables are described with mathematical equations. A *statistical model* is a mathematical model where some of the states or variables are regarded as stochastic. In a statistical model there exist variables that are known (or directly observed, given or assumed) and variables that are *hidden*.

In many cases it is convenient to speak of a *direction* in the model with internal states of the model influenced by an *input* (or “independent variables”, predictor, target, regressor, explanatory variable) and responding with an *output* (or “dependent variables”, response, regressand, explained variables), see figure 3.1. The variables in the model that regulate

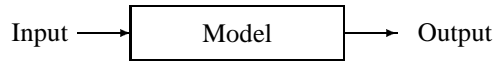


Figure 3.1: Model with input and output.

the input-output mapping is called *parameters*. Consider, e.g., the mathematical model

$$\mathbf{Y} = \mathbf{XB}, \quad (3.1)$$

where a data matrix \mathbf{X} can be considered the input which is multiplied by the model parameters \mathbf{B} to form the output \mathbf{Y} . This can be augmented to a statistical model by the inclusion of random disturbances \mathbf{U}

$$\mathbf{Y} = \mathbf{XB} + \mathbf{U}. \quad (3.2)$$

Now \mathbf{XB} forms hidden variables that are contaminated by the additive noise in \mathbf{U} before the model forms the output \mathbf{Y} .

The direction of the model is not always obvious, and in some cases it can be an advantage to do “inverse regression” where the direction is reversed (Krutchkoff 1967). Some analyses disregard the direction altogether viewing the modeling as symmetric. This has been called *interdependence analysis*, — as opposed to the symmetric *dependence analysis*, corresponding to what in the connectionistic literature has been called unsupervised and supervised modeling, respectively.

The operations that can be performed on the input-model-output system can be grouped according to what is known and unknown, see table 3.1. In prediction the input and model are known and the task is to predict an unknown output, e.g., in *classification* the prediction on the output is class labels and in *regression* the prediction is real values. Estimation concerns the identification of the model parameters either as point estimates or as (in Bayesian statistics) distributions. In reinforcement learning the output of the individual input samples are not known, but rather the effect of a set of input samples. Unsupervised learning can be described as a model that generates output. Inversion appears when the model and output is known. An example is PET reconstruction where the sinogram is the output, the model is obtained by the geometry of the scanner and the input is the site of positron annihilation (or better the beta decay). When there exists a linear relationship between the input and output the word “deconvolution” can be used. Actually since the convolution operator is “symmetric” (commutative) the estimation and inversion has the same mathematical formulation (Proakis and Manolakis 1996, section 4.6 “Inverse Systems and Deconvolution”), and, e.g., Boulanouar et al. (1996) use it in the sense of system identification: identify the model (the hemodynamic response function) from the input (the stimulus) and output (the fMRI signal).

Name	Input	Model	Output
Prediction	+	+	?
Estimation, supervised learning, system identification	+	?	+
Reinforcement learning	+	?	(+)
Unsupervised learning	NA	?	+
Inversion / deconvolution / inverse filtering	?	+	+
Blind deconvolution	?	?	+
Testing	+	+	+

Table 3.1: Operation on a model. “+” means that the component is known or measured, “(+)” only partly known, “?” that the component is unknown and is to be estimated. “NA” means that it does not apply.

Blind deconvolution tries to establish both the input and the model. It is possible to identify both the input and the model if some assumptions can be made about the data. Testing is where the discrepancy between the input, output and the model is assessed.

As an example process consider the input and output data to be known via a *training set*. This data are employed in estimation of the model parameters. Then the estimated model is used together with input and output from a *test set* to evaluate how well the model performs. If it performs well it is used to predict output from new inputs.

Below will the issues of estimation and testing be examined more closely.

3.2 Estimation

The words *training*, *learning*, *optimization*, *search* and *estimation* are used in several disciplines to denote the same operation, but weighing different aspect of it. They all denote the operation of finding a model parameter or a distribution of it. The word *search* is also used though mostly in combinatorial problems that do not incorporate any probabilistic variables in the model. *Learning* is often used in artificial intelligence (AI) to distinguish it from the classical AI where the “parameters” of the model are not estimated but set ad hoc by a human expert. Using the word *optimization* one usually wants to give the impression that the estimation is hard, e.g., nonlinear, requiring an iterative scheme. *Training* is often used in the neural network literature as a distinction between the other operations: validation and testing. *Optimization* will usually also mean that the estimation goal is a *point estimate*, — rather than a distribution of the parameter as in Bayesian estimation. *Estimation* is usually used when a probabilistic formulation has been made. In such a case one can think of estimation in the chain shown in figure 3.2. The first box of the chain contains a process where a statistical model

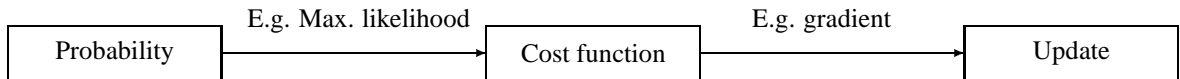


Figure 3.2: Chain for estimation of parameters in a model.

is setup. In the next box a cost function is derived by subjecting the statistical model to an estimation technique, and in the last stage an update rule is found that will optimize the cost function appropriately. Not all updates require a cost function and not all cost functions are developed from probabilistic assumptions. But even if the update is not developed from a cost function it might be associated implicitly with a one.

There are a number of different statistical estimation techniques/principals.

- The **maximum likelihood estimation** (MLE) principal maximizes the likelihood function $\mathcal{L}(\theta) = p(\mathbf{X}|\theta)$, often implemented indirectly by the minimization of the model parameters θ of the negative log-likelihood (Mardia et al. 1979, section 4.2)

$$\hat{\theta}_{\text{MLE}} = \arg \min_{\theta} [-\ell(\theta)] = \arg \min_{\theta} [-\log \mathcal{L}(\theta)] = \arg \min_{\theta} [-\log p(\mathbf{X}|\theta)], \quad (3.3)$$

where \mathbf{X} is the random variable of the statistical model. The negative log-likelihood is then regarded as the cost function.

- **Maximum a posteriori** (MAP) estimation is sometimes referred to as regularized maximum likelihood. Here the likelihood function is augmented with a prior on the parameters $p(\theta)$ (Ripley 1996, appendix A.1)

$$\hat{\theta}_{\text{MAP}} = \arg \max_{\theta} [p(\mathbf{X}|\theta) p(\theta)]. \quad (3.4)$$

When the logarithm is applied the prior will just become an additive penalty term to the log-likelihood. MAP has the advantage over MLE when there are many parameters Q compared to the number of objects N in \mathbf{X} : By biasing the estimate through the prior the estimate will be “better”, e.g., as assessed by the generalization error or the variance on the parameter estimate (Hoerl and Kennard 1970b), (Mardia et al. 1979, exercise 8.8.1).

- **Least (mean) square** aims at minimizing a least square error between a target (measured data \mathbf{Y}) and the model output/estimate ($\hat{\mathbf{Y}}$). In most cases this can be considered as a variation of maximum likelihood estimation with an implicit assumption of Gaussian error.
- **Minimum variance** methods aims at reducing the variance of the parameter estimate given assumption of the stochastic variables in the model. It often comes in conjunction with an assumption that the estimator should be unbiased, i.e., given $N \rightarrow \infty$ the parameter estimate should convergence to the true value. When the model is linear the optimal estimate is termed BLUE — best linear unbiased estimator.
- **M-estimation** is *robust* estimation suitable for data sets with outliers (Huber 1972). Often the methods can be reverse-engineered to MLE or MAP estimates with heavy-tailed distributions such as the Laplace distribution or adding a extra outlier class (Larsen, Andersen, Hintz-Madsen, and Hansen 1998a). The simple framework used in (Nielsen and Hansen 2001b) and section 5.9 where extreme points are discarded in a two-stage strategy can be considered as belonging to this domain and corresponds to an extended and multivariate approach in the line of α -trimmed means. See also (Mardia et al. 1979, section 3.2).
- **Bayesian inference** considers Bayes formula

$$p(\boldsymbol{\theta}|\mathbf{X}) = \frac{p(\mathbf{X}|\boldsymbol{\theta}) p(\boldsymbol{\theta})}{p(\mathbf{X})}, \quad (3.5)$$

where both \mathbf{X} and $\boldsymbol{\theta}$ are random. An unconditional so-called *prior* distribution for the parameters $p(\boldsymbol{\theta})$ is assumed as well as the conditional distribution (the likelihood) of the data $p(\mathbf{X}|\boldsymbol{\theta})$. Since the denominator can be found as

$$p(\mathbf{X}) = \int p(\mathbf{X}|\boldsymbol{\theta}) p(\boldsymbol{\theta}) d\boldsymbol{\theta}, \quad (3.6)$$

the *posterior* distribution $p(\boldsymbol{\theta}|\mathbf{X})$ can be found. In some simple cases a closed form expression can be found for the posterior. For complex models or distributions the posterior can only be evaluated approximately, e.g., with Markov chain Monte Carlo techniques (Neal 1996). Sometimes the full distribution is regarded as the result while if point estimates are desired the maximum or the mean can be used.

Below optimization and regularization are described in more detail.

3.2.1 Optimization

Optimization is not only considered as an estimation step in a statistical modeling process but appears in many other areas of mathematical modeling, and there are many forms optimization depending on the underlying problem. Some of the issues that determine what optimization type should be used are:

- **Constrained/unconstrained.** Some solutions require that the parameters of the model or elements of the model are restricted to have certain values. There are *hard* and *soft* constrains. Examples of soft constraints are that of ridge regression or weight decay of neural network where the parameters are encouraged to lie close to zero. In canonical correlation analysis the constraint that some of the latent variables should have the identity matrix as a covariance matrix can be considered. Soft constraints are often a result of regularization. Hard constrained optimization can often be transformed to unconstrained optimization by augmenting the cost function with *Lagrange multipliers*, see e.g. (Bishop 1995a, appendix C).
- **Continuous/discrete.** In continuous optimization the cost function is directly differentiable, and as many problems are *smooth* it is an advantage to utilize the derivative information, e.g., searching in the direction of the gradient.
- **Local/global.** In many optimization problems it is not necessary to find the global maximum, e.g., in a standard feed-forward artificial neural network there will be many “similar” global minima as the parameters can be permuted: The order of the hidden units can be change; the sign can be reversed between a layer and its next layer with affecting the value of the cost function (Bishop 1995a, section 4.4).
- **Noisy.** In some case the evaluation of the cost function cannot be made exactly. This might be because of numerical problems or because the cost function is stochastic, e.g., in connection with cross-validation with a limited validation set.

1st	2nd	Name	References
—	—	Amoeba / simplex / Nelder-Mead	(Nelder and Mead 1965), (Press et al. 1992, section 10.4)
+	—	Gradient descent / steepest decent	
+	—	Gradient descent with momentum	(Poljak 1964)
+	—	Quickprob	(Fahlman 1988)
+	—	Natural gradient	(Amari 1998)
+	—	Conjugate gradient algorithm	
		— Hestenes-Stiefel	(Hestenes and Stiefel 1952)
		— Fletcher-Reeves	(Fletcher and Reeves 1964)
		— Polak-Ribiere	(Polak and Ribiere 1969)
		— Scaled conjugate gradient	(Møller 1993)
+	—	Quasi-Newton — variable metric	
		— Powell-Broydon	
		— Davidson-Fletcher-Powell (DFP)	(Davidon 1959; Fletcher and Powell 1963)
		— Rank-one-formula / unirank	(Broyden 1967; Davidon 1968; Fiacco and McCormick 1968; Murtagh and Sargent 1969; Wolfe 1968)
		— Broyden-Fletcher-Goldfarb-Shanno (BFGS)	(Broyden 1970; Fletcher 1970; Goldfarb 1970; Shanno 1970)
+	(+)	Pseudo(-Gauss)-Newton	(Becker and Le Cun 1988; Ricotti, Ragazzani, and Martinelli 1988)
+	(+)	Gauss-Newton	
+	+	Levenberg-Marquardt	(Levenberg 1944; Marquardt 1963)
+	+	Newton(-Ralphson)	

Table 3.2: Unconstrained continuous variable multivariate local optimization techniques. “1st” and “2nd” column indicates whether the first and second order derivative is required as input to the algorithm.

- **Batch/on-line.** If the optimization is iterative the whole data set can be used (batch), a part of it (window), or a single object (stochastic, on-line, sequential, pattern-based).

A list of methods for continuous local optimization appear in table 3.2. Detailed descriptions of these methods are available in Fletcher (1980), Bishop (1995a, chapter 7), Ripley (1996, Appendix 5) and Hertz, Krogh, and Palmer (1991, section 6.2), — the last three with particular reference to optimization of neural networks. Some of the methods can be seen as using an approximation to the local Taylor expansion of the (cost) function f around the present parameter set θ_0

$$f(\theta) = f(\theta_0) + (\theta - \theta_0)^\top \underbrace{\frac{\partial f(\theta)}{\partial \theta} \Big|_{\theta_0}}_{\mathbf{g}_0} + \frac{1}{2}(\theta - \theta_0)^\top \underbrace{\frac{\partial^2 f(\theta)}{\partial \theta \partial \theta^\top} \Big|_{\theta_0}}_{\mathbf{H}_0} (\theta - \theta_0) + \dots \quad (3.7)$$

The Amoeba optimization (the top item in table 3.2) uses only the first term of the Taylor expansion, while the rest uses two of three terms — the first derivative (gradient, \mathbf{g}) and second derivative (Hessian, \mathbf{H}). The conjugate gradient and quasi-Newton methods construct the curvature information along the iteration steps and for the Gauss-Newton and Levenberg-Marquardt methods the second order derivative should be derived and feed as input to the algorithm. These last algorithms involve matrix inversion of the Hessian. The pseudo-Gauss-Newton method circumvents that by only considering the diagonal of the Hessian.

The iterative update formula for minimizing equation 3.7 with the Newton method is found by setting the gradient for the new parameter θ to zero

$$\min_{\theta} [f(\theta)] : \quad \mathbf{0} = \frac{\partial f(\theta)}{\partial \theta} \quad (3.8)$$

Third and higher order terms are ignored

$$\mathbf{0} = \mathbf{g}_0 + \mathbf{H}_0(\theta - \theta_0) \quad (3.9)$$

The guess for the new parameters θ is isolated

$$\theta = \theta_0 - \mathbf{H}_0^{-1} \mathbf{g}_0. \quad (3.10)$$

It is difficult to make generalization of the performance of the different algorithms over a general class of problems. The “No free lunch” theorem states that no algorithm is better than others when the performance on all cost functions is considered (Wolpert and Macready 1997). In practice with “normal” cost functions there are some aspects that can be generalized, e.g., of the 3 conjugate gradient algorithm it is my experience that the Hestenes-Stiefel does not perform well in connection with neural networks. It is often slower in convergence (for the cost function) than simple gradient descent. The Polak-Ribiere seems to be the best. Ripley (1996, page 345) and Bishop (1995a, page 281) also prefer this algorithm. The Lyngby toolbox implements several optimization methods for the artificial neural network, and the Levenberg-Marquardt algorithm is one of the fastest when the number of parameters that are optimized is small. Furthermore, it is quite *robust*, which is quite important when it has to “run on its own” where it, e.g., cannot be restarted if it jumps into a bad region and diverge. My experience is that even changes in seemingly non-important parameters as the multiplications factor in the Levenberg-Marquardt algorithm (the parameter that interpolate between the Gauss-Newton direction and the infinitely small gradient direction) can lead to significant different performance results of the optimization algorithm. This parameter can furthermore be determined in a number of different ways (Nielsen 1999).

Line search

Some of the multivariate optimization algorithm requires *line search* in each iteration, where the parameters are optimized along a given direction in the parameter space. This can either be hard/exact or soft/inexact/partial, with the former finding the minimum and the latter just a reduction in the cost function. There exist several methods: Bisection, Fibonacci, golden section, Newton, secant, parabolic and hyperbolic interpolation and “Brent’s algorithm” which is a combination of golden section and interpolation (Brent 1973). This step forms an extra complexity in the optimization.

3.2.2 Regularization and priors

When there are many parameters compared to the number of objects (examples) or if many of the input variables are highly statistical dependent it can be difficult to get a good parameter estimate. Manifestations are: The model *overfits* which means that the parameters are optimized to explain the particular data that has been used for the estimation (fitting both signal and noise) and it will perform poorly on a new set of data; a classical statistical test will *lose power* and typically not be able to reject a null hypothesis; the *variance of the parameter estimate will be large*. For some models, e.g., multilayer artificial neural network the variance of the parameter estimate will always be large since the parameters can be permuted.

The solution is to restrict the model or use related data. One of the classical methods is *ridge regression* that penalizes large parameter values in a linear model (Hoerl and Kennard 1970b). This is known as *shrinkage* or — in the neural network literature — as *weight decay* (Hinton 1986). Many regularization techniques can be developed from probabilistic assumptions with priors on the parameters, e.g., ridge regression estimate can be derived as a MAP estimate with a univariate Gaussian prior given for each parameter (Lindley and Smith 1972). Other distributions can be plugged into the model, e.g., Laplace distributions (Williams 1995; Goutte and Hansen 1997; Tibshirani 1996), generalized Gaussian (Fu 1998) or a general multivariate Gaussian distribution (Goutte, Nielsen, and Hansen 2000). The ridge regularization (and generalizations of it) were actually described earlier in connection with integral equations (Philips 1962; Tikhonov 1963; Tikhonov and Arsenin 1977) and a parallel line of literature to the statistical has been developed, see, e.g., (Hansen 1996) with ridge regression called *Tikhonov regularization*. Different types of regularization should sometimes be applied to the parameters, e.g., the parameter that models a constant term should not be shrunk (Brown 1977). Bishop (1995b) has shown that adding noise on the input is (under some assumptions) approximately equivalent to ridge regularization: One should refrain from adding noise to the data as more data points are generated, thus larger amounts of data has to be analyzed. Ridge regularization does not generate more data points, and the problem with setting the regularization parameters is also present in the case when adding noise to the data.

Another kind of regularization is *pruning*, that completely removes unimportant parameters from the model based on the effect they have on the cost function. *Optimal brain damage* (OBD, Le Cun, Denker, and Solla 1990) and *optimal brain surgeon* (OBS, Hassibi, Stock, and Wolff 1992; Hassibi and Stock 1993) are two methods introduced in the artificial neural network literature. Both algorithms rely on a Taylor expansion of the cost function. When the model is fully optimized the first order derivative will be zero, and the effect on the cost function when a parameter is deleted is assessed only by the second order term. OBD uses a diagonal approximation to the Hessian, while OBS will use the full Hessian.

In (Hansen and Rasmussen 1994) it is shown that an infinite weight decay on a parameter corresponds to pruning that parameter. Another combination of pruning and weight decay is *principal component pruning* (PCP) where activation of the units in successive layers is principal component analyzed and the small components are pruned (Levin, Leen, and Moody 1994). For a linear model this corresponds to principal component regression (Massy 1965), see also (Mardia, Kent, and Bibby 1979, section 8.8) or (Jackson 1991, section 12.3).

The setting of the hyperparameter might be critical in some applications others not. In the non-critical applications it is enough to set the regularization parameter on a value that is known to do perform reasonable well. In the critical applications one has to perform a tuning of the hyperparameter, — sometimes called “adaptive regularization”. This can

be done by cross-validation or algebraic estimates of the generalization error, e.g., adaptive regularization for regression neural networks (Hansen, Rasmussen, Svarer, and Larsen 1994; Hansen and Rasmussen 1994), and adaptive regularization for neural classifiers (Andersen, Larsen, Hansen, and Hintz-Madsen 1997), see also (Larsen, Svarer, Andersen, and Hansen 1998b; Bengio 2000). In many cases it is only feasible to give heuristic values for the optimization of the hyperparameter. The spatial normalization in SPM99 (Ashburner and Friston 1999) regularizes the deformation field. The tuning of the regularization parameter is based not on a quantitative measure but rather on a subjective judgment by the user (researcher) who should increase the amount of regularization “if [the] normalized images appear distorted”.

In the Lyngby toolbox simple regularization optimization is implemented in connection for the artificial neural network model for pruning size and ridge parameters using split-set validation set. This optimization is very time consuming for functional neuroimaging data given that each optimization of the model parameters for a given set of hyperparameters is non-trivial and already time consuming.

3.2.3 Non-orthogonality of design: Correlation between covariates

Optimally, the functional neuroimaging experiment should be designed so that variables of interest are applied independently. Unfortunately it is not always possible to construct such experiments. In certain experiments a subtask by default follows another task, e.g., a motion execution period is preceded by a motion preparation (Purushotham, Nielsen, Hansen, and Kim 1999).

One of the methods to circumvent this problem is by varying the period of the task. Purushotham, Nielsen, Hansen, and Kim (1999) varied the motion preparation. It was not possible to vary the motion execution as it was “controlled” by the subject. For fMRI it is necessary that the variation is large compare to the hemodynamic response — a few seconds might be enough.

3.3 Testing

“Testing” is the act of determining how probable a model is. “Validation” is a somewhat similar to testing the customary distinction between the two being that validation is performed as part of the parameter and hyperparameter estimation process, while testing is the assessment of the “final” model, see, e.g. (Ripley 1996, page 354).

There are several ways of testing:

- Classical statistical testing. One such approach is likelihood ratio testing (LRT) where two hypotheses are stated, i.e., the parameters of the model are fixed to some values or maximized within a region, and the resulting likelihoods are calculated and divided by each other, — the result being the *test statistics*. When distributional assumptions are made about the stochastic variables of the model the distribution of the test statistics can often be found analytically. The way in which LRT is usually employed is by fixing parameters to zero for one hypothesis and set the parameters for the other hypothesis so that the likelihood is maximized in the region different from zero.
- Bayesian tests. These tests resemble LRT when Bayes formula is used and priors are used as they are usually formulated as a ration between two probabilities with different hypotheses: *Posterior odds* and *Bayes factor*. The Posterior odds consider the ratio between the posterior of two models \mathcal{M}_1 and \mathcal{M}_2 (Ripley 1996, page 62)

$$\frac{p(\mathcal{M}_1|\mathbf{X})}{p(\mathcal{M}_2|\mathbf{X})}. \quad (3.11)$$

Considering the Bayes formula for one of the model \mathcal{M}_1

$$p(\mathcal{M}_1|\mathbf{X}) = \frac{p(\mathbf{X}|\mathcal{M}_1) p(\mathcal{M}_1)}{p(\mathbf{X})} \quad (3.12)$$

it is apparent that the term $p(\mathbf{X})$ is similar for both models. Furthermore, as the two models are considered discrete and in many case we want to assign equal prior on them $P(\mathcal{M}_1) = P(\mathcal{M}_2)$ — the *prior odds* are one, the posterior odds becomes equal to the likelihood ratio for the models, the so-called *Bayes factor* or “evidence ratio”

$$\frac{p(\mathcal{M}_1|\mathbf{X})}{p(\mathcal{M}_2|\mathbf{X})} = \frac{p(\mathbf{X}|\mathcal{M}_1)}{p(\mathbf{X}|\mathcal{M}_2)}, \quad (3.13)$$

with the evidence found by integrating the parameters in the model

$$p(\mathbf{X}|\mathcal{M}_1) = \int p(\mathbf{X}|\mathcal{M}_1, \boldsymbol{\theta}) p(\boldsymbol{\theta}|\mathcal{M}_1) d\boldsymbol{\theta}. \quad (3.14)$$

- Test-set based methods. Here the data is split and a part of the data is used for estimation (the *training set*), another part might be used to estimate hyperparameters (the *validation set*). The rest (the *test set*) is used for testing. When the objects are divided into v non-overlapping sets and validation is based on v different splits of training and validation set the method is referred to as *cross-validation*, — or v -fold cross-validation (Stone 1974). When there is only one object in the validation set it is termed *leave-one-out* (LOO) cross-validation. If the test set is picked at random the cross-validation is referred to as *Monte Carlo cross-validation* or “repeated learning testing” (Smyth 1996; Smyth 2000). Cross-validation with 50%/50% split are referred to as half-sampling (Celeux 2001). The approaches make little assumption of the underlying distributions. There are however several problems with it:
 - Some of the data cannot be used for the estimation of the model parameters, and if there is very little data this could mean that the parameters are not well determined.
 - A limited test set will have a variance compared to an infinite test set.
 - Test set based methods cannot directly deal with models that require fitting of parameters for the test data, such as in ANCOVA.

It is not clear what the optimal size of the training and validation set should be: Goutte and Larsen (1998) write that small split ratios (many training examples, few validation examples) tend to be best, and report accounts that find opposite results (Shao 1993; Larsen and Hansen 1995). Celeux (2001) finds the LOO outperforms half-sampling on small data sets while the opposite is the case for moderate to large sample sizes.

The model performance estimated on an (theoretical) infinite test set is called *generalization error*.

- Algebraic estimates of performance. These estimates works by using the error on the training set and augmenting it with a term that represents the overfitting. The term is typically a function of the number of parameters and the number of objects. Some of the many proposals in this domain are listed in table 3.3. With $\ell = \ln p(\mathbf{X}|\boldsymbol{\theta})$ as the log-likelihood (of the total data) and P and N as the number of parameters and objects some of them are:

$$\text{AIC} \propto \ell - P \quad (3.15)$$

$$\text{BIC} \propto \ell - P/2 \ln N \quad (3.16)$$

These have different characteristics, e.g., according to Hannan and Quinn (1979), Shibata (1976) has investigated the asymptotic properties of AIC: The estimate is not consistent, i.e., it asymptotically overestimates the number of parameters for autoregressive models with non-zero probabilities.

The difference in performance of the different methods has been assessed by several, e.g., the over-estimation of AIC is also found by Hansen, Larsen, and Kolenda (2001a) on simulated data analysed with ICA. Furthermore, it was found that BIC performed the best in comparison with test set and BIC on a simple simulated data set. On the other hand Smyth (1996) finds that BIC compared slightly poorer than cross-validation. This is also found by Celeux (2001) for poor sample sizes: “LOO outperforms BIC and ICL for poor sample sizes”. (Smyth 2000) compared cross-validation and BIC for clustering model selection on two data-sets (diabetes, atmospheric geopotential height) — the two methods returning the same result (3 clusters for each data set). Contrary, differences between ICL, BIC and AIC were found by (Goutte, Hansen, Liptrot, and Rostrup 2001) in assessment of the number of clusters in feature-space clustering of fMRI data.

3.3.1 The number of samples and learning curve

The number of samples (objects/scans/subjects) has an effect on the test. In classical statistical test the critical regions are often determined so the rate of false positive is controlled. That means that for few samples (e.g., subjects) the test will have small power not being able to detect a complex hypothesis, — and must choose the simple hypothesis. The issue has been addressed in functional neuroimaging by Friston, Holmes, and Worsley (1999a), and one of the recommendation in this connection is, e.g., that fixed effect analysis and conjunction analysis should be chosen for 6–10 subjects and random effects for above 12 subjects for fMRI studies.

In the artificial neural network literature the issue is often discussed as the *bias-variance trade-off* (Geman, Bienenstock, and Doursat 1992; Bishop 1995a) and *learning curves*. A learning curve is a tool to examine the relationship between the performance of an optimized model and the number of examples (scans): The generalization error is simply plotted against the size of the training and validation set. Often simple models, such as heavily regularized models or models only using a few components from a subset selection, are better than complex models on a small sample set because the complex models tend to overfit. With larger sample size the complex models usually perform better than the simple models, as the simple models is not able to model the complexity of the data. Thus it is often seen that the learning curves crosses (Mørch et al. 1997). The learning curve can also tell if new scans would provide better results: If the learning curve continues to decrease the model can be fitted yet better with new scans. However, if it has reached a plateau then it is likely that new scans would not help to get a model that fits the data better.

Name	Description	Reference
AIC	Akaike's Information criteria	(Akaike 1973; Akaike 1974)
AICC	Modified AIC	(Hurvich and Tsai 1989)
BIC	"B" information criterion	(Akaike 1977; Akaike 1978)
BIC/SBC	Bayesian information criterion	(Rissanen 1978; Schwarz 1978)
FPE	Final prediction error	(Akaike 1969; Akaike 1970)
FPE δ	FPE with an heuristic adjustment	(Bhansali and Downham 1977)
FPEC	FPE + test	(De Luna 1998)
GPE	Generalized prediction error, regularization FPE	(Moody 1991; Moody 1992)
FPER		(Larsen and Hansen 1994)
HQ		(Hannan and Quinn 1979)
ICL	Integrated completed likelihood	(Biernacki, Celeux, and Govaert 2000)
NIC	Network information criterion	(Murata et al. 1991; Murata et al. 1993; Murata et al. 1994)
RIC	Risk inflation criterion	(Foster and George 1994)

Table 3.3: Model order determination and generalization criterias. Akaike's BIC (information criterion B) is not the same as Rissanen's and Schwarz's BIC criterion, and to distinguish the two the latter has also been termed "Schwarz Bayes Criterion": SBC. See also (Ripley 1996, section 2.6). Several other criteria is described and evaluated in (Celeux, Biernacki, and Govaert 2001).

Kjems et al. (2000) using learning curves over subjects and multivariate discriminant analysis found that the model performance as a function of the number of subjects was depended upon the task: Simple tasks — such as figure tracing — were well described by 4 subjects while complex task — such as mirror tracing — require still more than 18 subjects.

The "net utility curve" of Genovese, Noll, and Eddy (1997, figure 4b) is a related technique.

3.3.2 Correlated and heterogeneous residuals

A problem for all kinds of tests appears if the samples (objects) are correlated, i.e., if the residuals are correlated across samples. If the sample represents a time-series the correlated residuals are often referred to as *colored noise*, — as opposed to white noise for uncorrelated residuals. A related problem appears when the residuals have inhomogeneous variance across samples. Collectively the two issues are referred to as non-sphericity of the residuals (Glaser et al. 2001).

In the case with test set and validation set based testing it is common to resample across sets of samples that are not correlated, e.g., in connection with multisubject fMRI and PET resampling should be done so that scans from the same subjects are not included both in the training and the test set. Such a scheme will only address serial correlation — not heterogeneous variance.

For classical test serial correlation in the residuals can be handled by correcting for the degrees of freedom resulting in *effective degrees of freedom*. Some early references are (Geisser and Greenhouse 1958; Box 1954).

A related correction appears if there are fewer *effective parameters* than the total number, e.g., because of regularization. In connection with artificial neural networks Moody (1991, 1992) has presented the "generalized prediction error" (GPE) as an extension to the "final prediction error" (FPE) (Akaike 1969) when the parameters of the model are regularized.

Discussions on the correction for the degrees of freedom in connection with functional neuroimaging data appear in (Ollinger and McAvoy 2000) and (Glaser et al. 2001). Modeling of the serial correlation in fMRI first appeared in (Friston et al. 1995c) and was later corrected in (Worsley and Friston 1995), the latter being identical to the Greenhouse-Geisser test (Glaser et al. 2001).

3.3.3 Simultaneous inference

When testing many hypotheses a special problem arises, e.g., in connection with mass-univariate testing in functional neuroimaging: Usually test thresholds are constructed so that in one out of twenty a true hypothesis will be rejected, e.g., "activation" will be attributed to a "non-activated" voxel. So if 20000 voxels are tested simultaneously then 1000 voxels will be categorized as being "activated" even though they are not activated. Table 3.4 shows some of the approaches to cope with this problem. Application of random fields in functional neuroimaging regards the residuals of neighboring voxels of the image volume as correlated as in a smooth Gaussian random field. Approximative distributions of various test statistics can be derived, such that p-values can be generated on height or extent (Cao and Worsley 2001). When drawing inference in random fields it is important to estimate the smoothness of the field correctly: The p-value will vary due to the smoothness estimate (Poline et al. 1997, section 7.4, Poline et al. 1995). The random fields corrections are

Method	Description	Reference
Bonferroni	Divide by the number of statistical tests	e.g. Armitage and Berry (1994, page 331)
FDR	False discovery rate	Benjamini and Hochberg (1995)
Random field	Gaussian	Worsley et al. (1992), Friston et al. (1991, 1994b)
Random field	Stationary resels; χ^2 , F , t	Worsley (1994)
Random field	T^2	Cao and Worsley (1999)
Random field	Gaussian scale space	Worsley et al. (1999)
	Scale space search	Poline and Mazoyer (1994)
CS	Null hypothesis accessed through simulated images	Ledberg et al. (1998)
MCS	Null hypothesis accessed through simulated images	Ledberg (2000)
Monte Carlo		Ward (1997)
Gaussian mixture	Gaussian mixture model in the summary image	Hartvig (1999)
Resampling		Turkheimer et al. (2000)

Table 3.4: Simultaneous inference: Compensation for multiple statistical comparison. For random fields see also the overview in (Cao and Worsley 2001).

implemented in the SPM toolbox and widely applied when reporting significant changes in PET and fMRI functional neuroimaging analyses.

Other descriptions of methods for simultaneous inference are (Forman, Cohen, Fitzgerald, Eddy, Mintun, and Noll 1995), (Xiong, Gao, Lancaster, and Fox 1995) and (Ward 1997).

An entirely different approach to simultaneous inference is by first doing a explorative mass-univariate voxel-based analysis (or any other type of explorative analysis) and establish one single specific hypothesis, e.g., about the activation in a region. Then this new hypothesis is tested on a new data set independent of the first data set used for exploration. (Friston 1997c) proposes a test if the region of the hypothesis is not known exactly.

None of the present techniques are implemented in the Lyngby toolbox. However, if the two-stage approach is taken the first stage need not consider a particular precise test, — and would not even need the computation of p-values but can return a region found to be interesting that is used in the second stage, — the region being based, e.g., on a threshold in a test statistics of a regressor or a saliency map.

3.4 Functional neuroimaging analysis

Modeling in functional neuroimaging is divided into preprocessing and analysis: The difference between the two is that preprocessing models the variables of no interests and analysis models the variables of interests. Sometimes it can, however, be desirable to model the variables of interest and of no interest simultaneously, thus blurring the border between preprocessing and analysis.

The goal of functional neuroimaging analysis is often to take a set of images (or image volumes) and process/analyze them in order to construct an image — often termed *summary image* — that as directly as possible relates to the behavioral component under investigation. In many cases it is appropriate to further characterize the summary image with a set of points, that describe maxima or center of local regions.

The Matlab program “SPM” (statistical parametric mapping) has been widely used for functional neuroimaging analysis and a general introduction to the methods and principles that this program rely on is available in (Frackowiak, Friston, Frith, Dolan, and Mazziotta 1997). The text in this book is more or less the same as in the “SPM course notes” (Institute of Neurology 1997).

3.5 Preprocessing in functional neuroimaging

Preprocessing in multisubject scientific functional neuroimaging studies is often necessary. A list of the preprocessing step are shown in table 3.5. The step in preprocessing are usually done one-by-one on a volume-by-volume basis. All these step should probably be done simultaneously for best results — potential they should be done simultaneously with the analysis — optimizing for best performance. However, at present this is not possible.

A review of preprocessing is available in (Krugel, Decombes, and von Cramon 1998). Maintz and Viergever (1998) have classified the different registration algorithms. A review of general medical image registration is available in (Bro-Nielsen 1996). Apart from validation on each single preprocessing step there are a few studies that try to determine the

PET	fMRI	Step	Description
✓	✓	Stripping	Remove the non-brain part of the image
	✓	Slice timing correction	Correct for different acquisition times in fMRI.
	✓	Inhomogeneity correction	Correct for scanner errors
✓	✓	Motion Correction	Correct for subjects movement
✓	✓	Coregistration	Alignment of different modality images.
✓	✓	Intersubject or atlas warping	Align and warping of subject to common frame
	✓	Removal of physiological noise	Removal of cardiac a respiratory effects
✓	(✓)	Normalization	Account for different doses
		Smoothing	

Table 3.5: Preprocessing steps or issues that has to be considered in functional neuroimaging studies (PET and fMRI).

effect of variation in several preprocessing parameters with Hopfinger, Büchel, Holmes, and Friston (2000) assessing change in resampling, temporal and spatial smoothing with z-score and p-values, and Kjems et al. (2000) that assess the preprocessing effect by multivariate analysis.

3.5.1 Reconstruction

The raw data measured by PET and fMRI scanners are not in a volumetric format. The PET raw data is returned as a *sinogram* and fMRI is returned in the Fourier domain also called the *k-space*. The transformation to the volumetric format is handled internally by the brain scanners, but the raw data is usually also available. There are different methods of reconstruction (Toft 1996) and they can affect the result of the functional neuroimaging analysis, e.g., in PET reconstruction a filter width needs to be set and, furthermore, the filtering can also be based on Markov random field reconstruction (Philipsen and Hansen 1999). Analysis directly on the PET sinogram data has been done by (Matthews et al. 1995).

3.5.2 Stripping

It is sometimes necessary to mask out the brain (*scalp-edited*), i.e., remove all non-brain voxels from the volume. There can be three reasons for that:

- Some voxels contains no part of the brain but are affected by the stimulus, e.g., in a saccadic eye movement study Law et al. (1998) found an artifact arising from increased (eye) muscle blood flow. (However, in this specific study the removal artifact was not removed, but explained post hoc.)
- For computational reasons it will be faster to exclude some voxels. Usually one ends with a fourth of the original voxels when discarding no-brain voxels. This will speed up the analysis by a fourth or more.
- Alignment software sometimes requires or benefits from masked brains (Kiebel, Ashburner, Poline, and Friston 1997).
- Anonymization of subjects if volumes are distributed from public databases.

Stripping MRI and fMRI is usually done is by constructing the histogram of the gray values and then apply a threshold at an appropriate value to obtain a binary volume. This binary volume is then put through morphological operations. Manual editing might be needed afterward. Other methods are based on surface description.

There are a few public programs for extraction of the brain: BSE (Sandor and Leahy 1997, <http://biron.usc.edu/~shattuck/bse/>), “Brain Extraction Tool” (BET) based on a surface model (Smith 2000) and mri3dX by Krish Singh (<http://www.mariarc.liv.ac.uk/mri3dX/>).

3.5.3 fMRI: Inhomogeneity correction, antialiasing, magnetization start-up

Inhomogeneity correction is desirable in connection with warping and segmentation, e.g., SPM99 uses intensity correction in connection with MRI segmentation (Ashburner and Friston 1997). An evaluation of the different inhomogeneity tools is available in (Arnold et al. 2001; Schaper et al. 2001).

For the fMRI scanner there is a transient magnetization start-up artifact causing the fMRI signal to have quite a different level than in the steady-state. There seems to be no reason to model this start-up effect, instead the scans are simply discarded.

3.5.4 Motion correction

Motion correction, *realignment* or *registration* aim at correcting for the subject head motion. Often the head of the subject is fixated, but it is not enough to avoid head motion. If there are plenty of scans or new ones can easily be acquired, then the scans with large motion artifacts can be discarded. Motion correction is usually corrected using a rigid body transformation. However, in connection with EPI MRI sequences it can be appropriate to do nonlinear transformations (Andersson 2001a; Andersson 2001b) due to the nonlinear spatial distortion in the MRI images.

Among the tools for motion correction are AIR (Woods et al. 1998a) and the functions in the different versions of SPM (Friston et al. 1995a; Friston et al. 1996d). “Unwarp” of Andersson (2001b) includes non-linear motion correction.

Motion correction is a time consuming preprocessing step as transformation parameters has to be estimated for each individual scan. Munk (1999) uses a neural network to speed up the registration algorithm.

An entirely different method of reducing motion artifact is “prospective acquisition correction” (PACE) where movement parameters are computed real-time during scanning and feed back to the sequence controller of the MRI scanner. This method can be combined with the ordinary (“retrospective”) motion correction (Erb, Hülsmann, Klose, Thesen, and Grodd 2001).

3.5.5 Coregistration — intermodality image registration

Coregistration, *intrasubject image registration*, *intermodality image registration* or *cross-modality registration* aim at aligning images from different modalities, such as PET and MRI, for the same subject, e.g., to map a structural image to a functional or vice versa. As with motion corrections usually only rigid body transformations are considered. However, EPI MRI images can require non-rigid body transformation (see section 3.5.4) and post mortem aMRI and stained histological material will also necessitate warping, being applied in, e.g., (Johnsrude, Cusack, Morosan, Halland, Brett, Zilles, and Frackowiak 2001). The estimation process is more complicated than the estimation of motion correction as the gray levels of the images cannot directly be compared. Among the different techniques are: Definitions of landmarks, alignment of segmented images, and alignment with a “flexible” cost function, such as, mutual information. A special problem is spectroscopic MRI and PET-ligand images where only signals from parts of the brain are present.

As coregistration is important in general radiology — and not just in functional neuroimaging — there exists a large number of algorithms, some of them shown in table 3.6. The most used are probably versions of AIR and SPM.

Name	Reference
AIR 1.0	(Woods et al. 1993)
FLIRT	(Jenkinson and Smith 2001)
SPM95	(Friston et al. 1995a)
SPM99	(Ashburner and Friston 1997), (Ashburner et al. 1997), (Maes et al. 1997)
—	(Andersson et al. 1995)
—	(Ardekani et al. 1995)
“Head and hat”	(Pelizzari et al. 1989)
“Ration image uniformity” (RIU)	(Woods et al. 1992)

Table 3.6: Coregistration algorithms.

Comparisons of different coregistration algorithms are available in (Strother et al. 1994; Turkington et al. 1993) involving a brain phantoms, (Kiebel et al. 1997) involving a simulated PET image. West et al. (1997) compared several image registration algorithm against a marker-based registration method.

3.5.6 Spatial normalization

Other names for the spatial normalization processing step are: *Normalization*, *spatial deformation*, (*intersubject image*) *registration*, *intersubject warping* and *atlas warping*, and it is a spatial transformation that is performed for two reasons:

- In a multisubject study: Transform volumes acquired from different subjects to the same space. This transformation accounts for intersubject variability in gyral anatomy.
- Transform volumes to a standardized space so that coordinates can be compared across studies.

The result is affected by the algorithm and by the template chosen. Table 3.7 lists some of the many algorithms in use and table 3.8 some brain templates. In many cases the template is tied to an algorithm. Special templates can be necessary for children and ligand PET images, and images containing abnormal deformities, and may also require special algorithms, e.g., large regularization in the warp or small number of parameters (Stamatakis, Wilson, and Wyper 2000).

Name	Description	Reference
AIR3		Woods et al. (1998b, Woods et al. (1999)
CHSP		San Antonio
MRIWarp	Non-linear warp	Kjems et al. (1999)
SN	9-parameter affine transformation	Lancaster et al. (1995)
SPM95		Friston et al. (1995a)
SPM96	$4 \times 5 \times 4$ basis functions	Ashburner and Friston (1996)
SPM99	Default is a $7 \times 8 \times 7$ basis function	Friston et al. (1995a), Ashburner et al. (1997), Ashburner and Friston (1999)
STAR	Elastic warping	Davatzikos (1997)
—	Stockholm	Greitz et al. (1991) Ingvar et al. (1994)

Table 3.7: Spatial normalization algorithms and software.

Name	Description	Reference
SPM96	MNI single subject (Colin Holmes). Also used in BrainWeb, in the same space as MNI305	SPM99 spm_templates.man
MNI152	Standard template in SPM99, distributed volume are smooth with 8mm FWHM	SPM99 spm_templates.man
MNI305	ICBM standard, also distributed in SPM99	SPM99 spm_templates.man
Visible Human	Brain from the Visible Human Project	http://www.nlm.nih.gov/research/visible/visible_human.html
VAPET	Used at the VA Medical Center, Minneapolis	
CBA	“Computerized brain atlas”, Dept. Neuroradiology, Karolinska Institute	(Greitz et al. 1991), (Seitz et al. 1990)
HBA	“Human Brain Atlas” from Karolinska Institutet	(Roland et al. 1994)
new HBA	Re-acquired HBA used in European Computerised Human Brain Database	(Schormann et al. 1999; Roland et al. 1999)
“BIT”	Warped single subject	(Lancaster et al. 2001)
EVA833	Template based on 833 elderly subjects	(Quinton et al. 1999)
—	Ligand templates	(Meyer et al. 1999)
Talairach	Original Talairach images. No MRI exist and it has never been used directly as template.	(Talairach and Tournoux 1988)

Table 3.8: Templates: Some of the standard brains used to atlas warping

Test and validation of the spatial normalization is difficult. There are basically two methods: One method is to measure the discrepancy in manually defined landmarks; the other is to put the resulting transformed volumes through statistical analysis and compare the significance values. Davatzikos et al. (2001) compared their STAR algorithm against SPM95, SPM96 and SPM99 and found their method better than SPM, with the following ranking: STAR - SPM99 - SPM96 - SPM95.

Many functional brain mapping studies are reporting their findings in the “Talairach space”. This space — or atlas — is defined in a book of Talairach and Tournoux (1988) where a segmented brain is displayed in detailed color plates. There does not exist an MRI scan of the subject in the Talairach book and as the warping algorithm usually requires these scans for the registration the Talairach atlas cannot be used directly.

The International Consortium For Brain Mapping (ICBM) has developed a number of brain atlases. Their official standard is the so-called MNI305 based on the average of 305 stereotactic T1-weighted volumes.¹ Averaging over such a large number of subjects makes the image very blurred and except for the large fissures none of the sulci are distinguishable. The MNI305 template is distributed with SPM96 in a $2 \times 2 \times 2$ mm format. Apart from being a spatial normalization template the ICBM atlas is also a probabilistic atlas, defining probability volume for, e.g., the cerebellum (Mazziotta et al. 1995).

In comparison and meta-analysis between studies using different templates it is of course important that the anatomical locations correspond. One of the few transformations is between the Talairach atlas and the MNI space: Andreas Meyer-

¹An web-based interface to the images is presently available at http://www.bic.mni.mcgill.ca/cgi/icbm_view

Name	Description	Reference
ANIMAL	“Automatic Non-linear Image Matching and Anatomical Labeling” Nonlinear warping and labeling by a previous labeled volume	(Collins, Holmes, Peters, and Evans 1995)
FAST	Segmentation with Markov random field	(Zhang et al. 2001)
INSECT	GM, WM and CSF segmentation with an artificial neural network with 9-parameter spatial normalization	(Kollokian 1996; Collins et al. 1994)
SEAL	“Sulcal Extraction and Automated Labelling”	(Goualher et al. 1999)
SPM99	Segments into GM, WM, CSF and other	(Ashburner and Friston 1997)
—	Cortex surface extraction	(MacDonald et al. 2000)
—	Combined manual/automatic	(Zavaljevski et al. 2000)

Table 3.9: Some of the segmentation algorithms and tools available.

Lindenberg has suggested an affine transformation (Meyer-Lindenberg 1998; Brett 1999) (the unit is meter):

$$x' = 0.88x - 0.0008 \quad (3.17)$$

$$y' = 0.97y - 0.00332 \quad (3.18)$$

$$z' = 0.05y + 0.88z - 0.00044 \quad (3.19)$$

A transformation using a standard voxel-based spatial normalization procedure was used between the MNI and an artificial constructed Talairach from the Talairach Daemon showed a closer match between the two templates especially in the inferior temporal lobe (Brett, Christoff, Cusack, and Lancaster 2001).

Apart from being a preprocessing step in activation studies, spatial normalization can also be used in morphometry where anatomical shape difference among groups are analyzed. An example of this is the deformation-based morphometry approach of Gaser et al. (2001) using the warp algorithm of Kjems et al. (1999).

The parameters of spatial normalization are an important set of parameters to record in coordinated-based neuroinformatics databases. Discrepancies between the brain templates will make it difficult to do meta-analyses on the coordinates, e.g., the estimation will be biased for meta-analysis on intersubject variability of activation. It is not always clear from the functional neuroimaging article precisely what template and what method has been used.

3.5.7 Segmentation

Segmentation usually classifies the voxels of the brain into white matter (WM), gray matter (GM) and cerebral spinal fluid (CSF), but sometimes other classes are used, e.g., white matter lesions as in (Garde, Mortensen, Krabbe, Rostrup, and Larsson 2000) or “cerebral cortex”. Among the purposes of segmentation are correction for partial volume effects, extraction of the brain (for visualization purposes or if other components the processing chain require it) and quantitative analysis of the amount GM or lesions, e.g., Garde, Mortensen, Krabbe, Rostrup, and Larsson (2000) relate white matter lesions with decay in intelligence, — here the white matter lesion segmentation was done manually by a radiologist. The segmentation is often performed with Gaussian mixture models on MRIs, — either just a single type or several: T1/T2/proton density. The segmentation algorithms might incorporate two kinds of priors: a prior based on the classification of a neighboring voxel or a prior on the specific stereotactical coordinate of the voxel. Inhomogeneity correction is often necessary.

Apart from Gaussian mixture models deformable models have also been applied in segmentation. One of the successful has been (MacDonald et al. 2000; Kabani et al. 2001), that fits the outer and inner surface of cortex simultaneously with constraints. This approach is also taken in SureFit (Van Essen, Drury, and Anderson 1999; Drury, Van Essen, and Anderson 2000).

Manual and automated segmentation can be combined by first doing manual segmentation on a small number of voxels, then do supervised learning with the label as targets and the voxel data as input, and finally predict the classification for the rest of the voxels with the estimated model (Zavaljevski, Dhawan, Gaskil, Ball, and Johnson 2000).

3.5.8 Confounds and nuisances in PET and fMRI: Removal of physiological noise, ...

Apart from head motion artifacts there is a large number of other confounds that influence the image data. These confounds can usually be grouped into either *physiological* or *instrumental/experimental*, they include among others cardiac and respiration confounds (Noll et al. 1996), the effect of swallowing, low frequency oscillations (Lowe et al. 2000), “Global effect” (Aguirre et al. 1998a), bolus size, scanner drift, susceptibility and ghosting.

Method	Reference
Manual selection of artifact voxels	(Petersen et al. 1998)
Alignment, variance and ROI-definition	(Lund and Hanson 2001)
Clustering, fuzzy	(Scarth et al. 1996)
ICA	(Beckmann et al. 2000), (Netsiri et al. 2000)

Table 3.10: Identification of artifacts in fMRI.

Method	Description	Reference
Linear drift	Removal of the linear drift	(Bandettini et al. 1993)
Linear drift	Removal of the linear drift estimated with base line scans	(Goutte et al. 1999a)
Quadratic detrending		(Cox 1996)
Discrete cosine set	Implemented in SPM99	(Holmes et al. 1997)
Prediction	Prediction of fMRI nuisance from physiological recordings	(Park et al. 2000)
Running-line smoother		(Marchini and Ripley 2000)
PCA	PCA on the residuals orthogonal to the design	(Ardekani et al. 1999)
“State space”	Random walk and sinusoidal with AR frequency	(Petersen et al. 1998), (Hartvig 2000, paper 6)

Table 3.11: Modeling and removal of artifacts

Some of the confounds yield low frequency noise — also called 1/f-noise or pink noise — in fMRI: It is usually attributed to long term physiological drift, head movement (even after realignment) (Zarahn, Aguirre, and D’Esposito 1997) and/or scanner instabilities (Smith, Lewis, Ruttiman, Ye, Sinnweell, Yang, Duyn, and Frank 1999).

Due to the relative large amount of noise in the lower parts of the spectrum in fMRI it is an advantage to design experiments where the paradigm is rapid, e.g., stimulation blocks should not be 1 minute long.

The approaches to model the confounds can be classified according to the data required to model the confounds. When the confound is identified it can either be used as a covariate (e.g., in a GLM model) or extracted prior to the modeling of the interesting signal. Required data:

- **None.** Use “fixed” models to remove the artifact. An example is filtering in fMRI, where a high pass filter removes the 1/f-noise. Such an approach is, e.g., taken by Holmes, Josephs, Büchel, and Friston (1997) that use a set of discrete cosine functions.
- **Image data.** Estimate the confound from the image data. There are several ways to identify the confounds, see table 3.10. The use of independent component analysis for this has been suggested by several. Lund and Hanson (2001) suggest a relatively automatic procedure for modeling of physiological noise by identifying vessel voxels by angiogram and BOLD standard deviation images.
- **Additional data.** Acquire additional measurements and use that in modeling the confounds. Examples are typically cardiac and respiratory measurements (Hu, Le, Parrish, and Erhard 1995).

Susceptibility artifacts arise especially in the brain-air interface, as is the case for “temporal lobes near the air-filled sinuses” where the fMRI shows loss of signal (Devlin et al. 2000a), (Devlin et al. 2000b). The effect of this can be assessed by comparing T2*-weighted images with T2 weighted images or by comparing the differences in signal intensities across the gray matter Lipschutz et al. (2000) Lipshutz et al. (2001), and different methods to compensate for this has been suggested applied either during the image acquisition (Cordes, Turski, and Sorensen 2000) or in the analysis stage (Devlin et al. 2000a), (Devlin et al. 2000b).

Ghosting occurs in the EPI sequences of fMRI: The image is shifted and may be overlayed onto the original. This is somewhat similar to normal aliasing in signal processing. It can be quite severe, e.g., Andersson (2000) found that a checkerboard stimulus would generate false positive in the frontal lobe. Clare (1997, section 4.3.1) describes some of the method to model the artifact. These methods are not applied in mainstream analysis, e.g., SPM does not model for ghosting.

Autocorrelation

The fMRI time-series will be autocorrelated either because the stimulus itself autocorrelated (e.g., as in block design) or through the hemodynamic response. But also other sources, such as physiological noise, can generate autocorrelation, see, e.g., (Purdon and Weisskoff 1998). This is related to the 1/f-noise mentioned in the previous section. Statistical

inference will be biased if the autocorrelation is not modeled (the DOF is overestimated) and there have been proposed several approaches for modeling autocorrelation:

- Modeling the noise by whitening (pre-whitening). Purdon and Weisskoff (1998) use an first order autoregressive (AR) model while Lange and Zeger (1997) employ a spatial smoothed estimate of the noise in intervals in the frequency domain.
- Smoothing (“precoloring” or swamping), where both the design matrix (the input function) and the fMRI data (the target) (the fMRI response are “over-smooth” with a specific kernel, (Friston, Holmes, Poline, Grasby, Williams, Frackowiak, and Turner 1995c; Worsley and Friston 1995). A Gaussian or a fixed hemodynamic response is chosen to construct the kernel. Purdon and Weisskoff (1998) referred to this as the *extended general linear model* (E-GLM), and if only the design matrix is smoothed they referred to it as *modified extended general linear model* (ME-GLM).

A discussion of these issues is available in (Purdon and Weisskoff 1998; Friston, Josephs, Zarahn, Holmes, Rouquette, and Poline 2000a).

It can be noted that Kershaw, Ardekani, and Kanno (1999) argue that there is no temporal correlation if the confounds are modeled well enough: The reason why temporal correlation is found in several studies is due to a too simple modeling of the confounding signals, e.g., a model with simple mean and linear drift removal “required” a first-order autoregressive model to account for the noise, while a more complex confound model (as the one used by Ardekani, Kershaw, Kashikura, and Kanno 1999) did not.

Modeling of the autocorrelation is not implemented in the Lyngby toolbox. However, the effect of it can be assessed by studying the (normalized) cumulated periodogram where the normalized cumulated sum of the magnitude of the Fourier transformed residuals are plotted against the frequency (Madsen 1989). If the residuals are white then the curve should follow a straight line. Figure 3.3 shows an example of a highly activated voxel from the data acquired in (Friston, Jezzard, and Turner 1994a) after a FIR filter modeling with the Lyngby toolbox. For this voxel the low frequencies are dominating (the curve is *over* the straight line) and this indicates that there is autocorrelation left in the residuals.

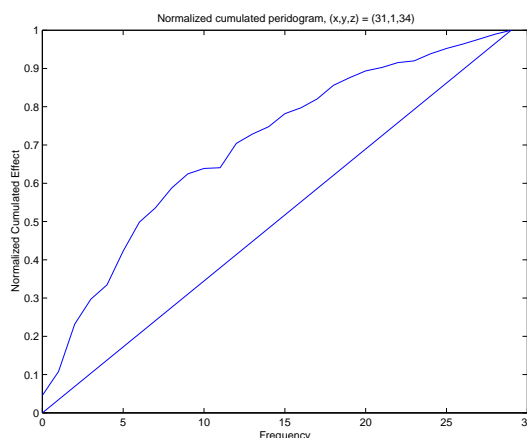


Figure 3.3: Normalized cumulated periodogram for the residuals of a highly activated voxel in the visual cortex with checkerboard stimulation from the study of Friston, Jezzard, and Turner (1994a) after a FIR filter modeling.

Normalization

In connection with PET studies the amount of tracer in the brain may vary, e.g., because of variation in the bolus size, and the effect of this has to be considered and possibly modeled, i.e., the effect of the global blood flow (gCBF) on the regional blood flow (rCBF). In some studies gCBF is measured for each scan by other means than PET (by arterial blood sampling) and the value so obtained can be used to scale the images. In other studies the normalization parameters are determined from the images, some of the methods being listed in table 3.12. ANCOVA and proportional scaling are the most common. ANCOVA normalization regards the effect of the gCBF on the individual voxels as additive, while “proportional” scaling regards it as multiplicative. It is not clear what is the most appropriate scaling (Andersson 1999b):

[...] if one believes that the dominating source of variance in the globals is ‘apparatus dependent’ (i.e. differences in injected dose (PET) or drifts in amplifier electronics (fMRI) then proportional scaling is clearly more correct. If on the other hand one believes that the dominating source is ‘physiological’ then the issue becomes less clear.

Name	Description	Reference
ANCOVA	Additive	Friston et al. (1990)
Proportional	Also called “ratio approach” or “scaling”	(Fox and Raichle 1984)
“Andersson”	Activated voxels discarded from estimation by iterative F-masking	Andersson (1997)
“Andersson2001”	Estimation of gCBF through optimization of stochastic sign changes	Andersson et al. (2001)
Histogram	Histogram equalization	Arndt et al. (1996)

Table 3.12: Normalization methods for gCBF in PET.

Another issue is to estimate the gCBF, e.g., it should probably be avoided to use the entire image as some voxels will not contain the brain and/or cortex. An approach taken by SPM is to find the mean of the voxels that are above an eighth of the mean of the entire image (Andersson 1999a).

A further complication is if the gCBF is correlated with the task. This seems to depend on the task (Law 1996, section 1.7.2.5), and can have an undesirable effect on the analysis (Aguirre, Zarahn, and D’Esposito 1998a).

3.5.9 Spatial Filtering

There are several reasons for applying spatial filtering to each of the volumes:

- Increase the signal-to-noise ratio. Due to the matched filter theorem it has often been recommended to use a large width filter when looking for activation in the cortex and small width filters when looking for activations in the small basal ganglia.
- Accommodate for the anatomical variation between subjects. For single-subject studies the filtering width can be smaller (e.g., FWHM $7 \times 7 \times 7$ mm) than for multi-subject studies (e.g., $10 \times 10 \times 10$ mm or more).
- Make the volume a Gaussian random field.

Spatial filtering is almost always done in multi-subject studies and is a normal procedure in SPM. The filtering method is usually done with either a boxcar (mean/median) or a Gaussian isotropic filtering. However, other suggestions have been Markovian spatial filtering (Descombes, Kruggel, and von Cramon 1998), atlas based adaptive filtering (Davatzikos, Li, Herskovits, and Resnick 2001), and multi-level filtering (Poline and Mazoyer 1994; Worsley, Marrett, Neelin, and Evans 1996). It has also been suggested to smooth *after* the analysis (Lowe and Sorensen 1997; Maisog 2000).

3.5.10 Reduction of the number of analyzed voxels

Before the actual analysis the number of voxels can be reduced by applying a simple conservative test. In SPM this is called “F-statistic threshold filtering” (Holmes and Friston 1997, section 6.4) or “F-masking”. There can be two reasons for doing this: Reduction the number of voxel will reduce the computational burden and the reduction can be an advantage in, e.g., cluster analysis where many non-interesting (non-activated) voxels might affect the clustering of interesting voxels (Goutte, Toft, Rostrup, Nielsen, and Hansen 1999b).

In the usually SPM setting the F-test takes up relatively many degrees of freedom and it is possible that a specific contrast might be significant while the F-test is not. To be on the safe side a conservative threshold for the F-test should be employed.

Another more heuristic approach was taken in, e.g., (Friston et al. 1995b, section “data processing”) where voxels below 0.8 of the volume mean were excluded, with the aim of restricting the analysis to the intracranial region. Somorjai, Jarmasz, Baumgartner, and Richter (1999) reduces the number of voxels by auto-correlation and trend test, and (Purushotham, Nielsen, Hansen, and Kim 1999) applied K-means clustering to identify interesting voxels for further analysis.

3.6 Analysis of functional neuroimages

A general overview of the analysis methods for functional neuroimaging is available in (Petersson, Nichols, Poline, and Holmes 1999a; Petersson, Nichols, Poline, and Holmes 1999b; Horwitz, Friston, and Taylor 2000). Reviews are available in (McColl, Holmes, and Ford 1994; Lange 1996). Among the first mathematical/statistical analyses of functional neuroimages were (Moeller, Strother, Sidtis, and Rottenberg 1987; Moeller and Strother 1991), (Fox and Mintun 1989) and (Friston, Frith, Liddle, and Frackowiak 1991). Below the analysis techniques are grouped into *unsupervised* and *supervised* algorithms.

Comparisons of the many different analyses are available in (Xiong, Gao, Lancaster, and Fox 1996) and (Lange, Strother, Anderson, Nielsen, Holmes, Kolenda, Savoy, and Hansen 1999; Lange, Hansen, Pedersen, Savoy, and Strother 1996; Lange, Hansen, Anderson, Nielsen, Savoy, Kim, and Strother 1998).

3.7 Unsupervised methods

Unsupervised modeling is modeling (of a set of observation that can be contain in a matrix $\mathbf{X}(N \times P)$) without (using) any target/design for the optimization of the model parameters. Among the methods that is usually regarded as unsupervised are singular value decomposition (SVD), principal component analysis (PCA), multidimensional scaling (MDS), self-organizing maps (SOM), non-negative matrix factorization (NMF), independent component analysis (ICA), cluster analysis (with fuzzy and K-means), kernel density estimation (KDE), and Gaussian mixture models (GMM). Most of these methods can be used in a probability density estimation. Many of the unsupervised methods reduce the data into some interpretable *components*. These components can be characterized in a number of ways:

- **“Direction” versus “point”**. The components can be associated with a direction in the space spanned by the data or an absolute point in the space. Cluster analysis would usually be associated with clustering data as points in the space, whereas PCA would consider the direction. However, cluster analysis can also be performed on the direction by suitable normalization.
- **Sparse versus dense coding**. Some of the algorithms produce what is called “sparse coding” or localized coding: \mathbf{X} is factorized and the obtained components contain many elements that are zero or close to zero in the case of sparse coding, i.e., the component are effected only by a restricted set of the P variables in \mathbf{X} . Examples of methods that can produce sparse coding are ICA and NMF. Lee and Seung (1999) give an example of NMF of a data matrix with faces. The obtained components (the basis images) contain several version of mouths, noses and other facial parts, whereas the VQ and PCA components are not restricted to a localized part of the image.
- **Orthogonality**. In (the draft to) Friston et al. (1996a) write about an MDS analysis of a functional neuroimaging study with word generation:

[...] each dimension is associated with a particular profile of the experimental conditions. For example the first dimension points in the direction of all the intrinsic word generation tasks and away from the baseline word shadowing tasks. Conversely the second dimension points towards the first scans and away from the last scans.

It need not necessarily be so. In a multisubject study the first principal component will usually describe variation among the subjects, and it seems not to be natural to require that the space of the interesting variable should be orthogonal to the space that models the intersubject variation. Both cluster analysis and ICA do not have that restriction.

- **Interpretability**. For reasons of human interpretability one would like to have as few components as possible. Methods such as kernel density estimators have as many components as objects and the “components” of the result will not be more interpretable than the original data.

Unsupervised methods in brain mapping have their advantage when the input (e.g. the stimulus) is not well defined or if there are variables that are not controllable by the experiment.

It is not correct to say that the unsupervised methods are not “statistically” methods, e.g. (Friston, Poline, Holmes, Frith, and Frackowiak 1996b) write about principal component analysis:

Although powerful, in a descriptive sense, eigenimage analysis and related approaches are not “statistical” methods that can be used to make inferences about the significance of the results; they are mathematical devices that simply identify important patterns of correlations of functional connectivity.

There *do exist* statistical tests for principal component analysis (eigenimage analysis), e.g., the isotropy test (Mardia, Kent, and Bibby 1979, section 8.8). What makes multivariate tests less related to the experimental hypothesis is:

- Multivariate analysis tests do not usually localize the test, so the inference is for the *whole image* rather than a voxel or a local region.
- Functional neuroimages are often dominated by confounds and nuisances, e.g., in a multi-subject study it is often found that the dominating variance is the between subject variance rather than the variance imposed by the different experimental conditions (Strother, Kanno, and Rottenberg 1995b; Strother, Sidtis, Anderson, Hansen, Schaper, and Rottenberg 1996b). An isotropy test in this connection would just test if “something happens” – not that this is of neuroscientific interest.

3.8 Cluster analysis

Cluster analysis assigns N observations with unknown classification into K groups (“clusters”). This can either be done so the assignment is *exclusively* to one group (hard threshold, e.g., K-means) or *fuzzy* where the observation is assigned to all groups with different weights, — a weight that can be regarded as a probability for belonging to a group (soft threshold, e.g., fuzzy clustering, Gaussian mixture). The observations in a group should be “homogeneous” in some measure. Often the observation within a group is assumed to be generated from the k ’th of K Gaussian distributions

$$p(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k, k) = |2\pi\boldsymbol{\Sigma}_k|^{-1/2} \exp \left[-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu}_k)^\top \boldsymbol{\Sigma}_k^{-1} (\mathbf{x} - \boldsymbol{\mu}_k) \right]. \quad (3.20)$$

If the prior for assignment is equally among the K clusters a hard threshold assignment for an observation \mathbf{x}_n can be obtained by choosing the k ’th that has the highest probability, — or equivalent: shortest Mahalanobis distance. Some of

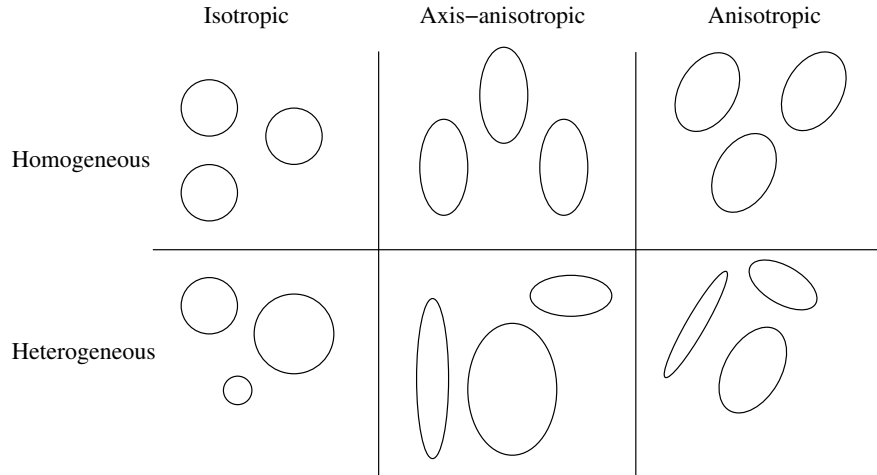


Figure 3.4: Cluster types for a Gaussian clusters (or another elliptic distribution). The ellipses denote isocurves for a threshold in 2D probability densities: Small ellipses has a “small” covariance. The first row depicts cases where the distribution is the same for all clusters and the second row where they are different.

the different forms of the covariance matrices $\boldsymbol{\Sigma}_k$ can be seen in figure 3.4, see also (Fraley and Raftery 1998, table 1) which list references for the different variation of covariance parametrization. Variation in assignment can also be produced by weighing the different clusters with different prior probabilities. Based on results from their own simulated data set (Celeux and Covaert 1995) advocate different priors (volumes) for each cluster.

The different clustering algorithm can be classified according to how the mean, covariance and assignment are estimated or assumed. If the assignment is assumed in advance then it is usually not called an unsupervised cluster analysis, but rather a supervised classification.

The K-means clustering algorithm (MacQueen 1967) assumes isotropic homogeneous covariances, and the means are estimated as the arithmetic means of the presently assigned data points to the cluster. The data points are assigned with hard threshold to the clusters. As the assignment changes when the means change, and the means change when the assignment change an iterative scheme is necessary where the mean and assignment are alternately estimated. A variation of the K-means algorithm is the K-mediod algorithm (Vinod 1969), where the means of the clusters are found as the median among the assigned data points.

Clustering algorithms with isotropic covariances will only be relevant if the variables are measured in the same unit: When the variables represent different scales, as e.g., in (Goutte, Nielsen, Svarer, Rostrup, and Hansen 1998) where three different features were extracted from the result of a cross-correlation analysis, a unique clustering can be obtained by scaling the variables with their standard deviations. Another issue with isotropic covariances is that the resulting clusters can have different magnitudes and not different forms, e.g., if the hemodynamic response function is modeled the different clusters might represent responses with different amplitudes rather than with different forms, such as delay. This issue can largely be resolved by scaling. Hartigan (1975, section 4.9.1) also makes a heuristic suggestion that after K-means clustering the intraclass covariance matrix is calculated and if it differs significantly from a multiple of the unit matrix the data should be transformed. Furthermore, according to Bottou and Bengio (1995, figure 1) it is preferable to use an online version of the K-means algorithm rather than a batch, though the online version is faster for the first few epochs.

There exist different clustering algorithm than K-means, e.g., fuzzy clustering (Bezdek 1981) and hierarchical clustering generally either divisive (“successive partitions of the sets of objects”) or agglomerative (“successive pooling of subsets of the set of objects”) (Mardia et al. 1979, section 13.3). The fuzzy clustering requires both the number of clusters

and a fuzziness index specified. If distributional assumption are made then mixture models can be used for clustering. This has been considered by several authors, e.g., (Banfield 1993; Celeux and Covaert 1995).

The means can be initialized by selecting random data points as the mean. This provides a good spread of the cluster centers in the data space. The Lyngby toolbox provides a few more cluster center initialization methods that are suitable when the paradigm is known. These are based on computation of the cross-correlation and choosing random data points spread out on the sorted magnitudes of the cross-correlation. (Christiansen 2000, section 5.4.1) discusses a few other initialization methods.

Clustering does not have to be performed in the original data space, but can also be performed in the space obtained from cross-correlation analysis (Goutte, Toft, Rostrup, Nielsen, and Hansen 1999b) or canonical correlation analysis (Akaho, Hayamizu, Hasegawa, Yoshimura, and Asoh 1997). K-means clustering on the cross-correlation function is implemented in the Lyngby toolbox.

3.8.1 Choosing the number of groups

Wiegell, Tuch, Larsson, and Wedeen (2000) clustered voxels of thalamic nuclei based on diffusion weighted images and the number of groups in the cluster analysis could be set to the known number of nuclei. However, in many cases where cluster analysis is applied the number of groups (clusters) is not known a priori. Often the number is set to a number the researcher finds “suitable”: To a number that produces easy interpretable results. This will often be a subjective judgment and it is better to use objective methods that only depend on the data: There does not exist a principled method to determine the number of groups in cluster analysis. There are a few heuristics and some developed from a probability density modeling view of cluster analysis. McLachlan (1982) and Everitt (1979) give reviews of the problem. Palubinskas et al. (1998) introduce a prior for the number of clusters which moves the problem from determining the number of clusters to determining the value of a hyperparameter. Some of the suggested methods have been:

- Sudden flattening in within-variance (Thorndike 1953). The first step of the algorithm is the computation of the within-variance for all clusterings up to a sufficiently large number of groups yielding a monotonically decreasing function of the within-variance as a function of the number of groups. The second step is to compute the curvature of the decreasing function and identifying a sudden change. Everitt (1979) describes that Gower (1975) used a similar method. In cluster analysis for functional neuroimaging the method has been used by Goutte et al. (1999b).
- $K^2|\mathbf{W}|$ suggested by Marriott (1971), see also (Mardia et al. 1979, page 365). This algorithm simply augments the within-variance with the squared number of groups and identify the value of the number of groups where the function has its minimum. It will be a flat curve for a uniform distribution.
- Kurtosis. If the clusters are assumed to be Gaussian a large kurtosis of a cluster indicates that the cluster should be described by two clusters (two Gaussian distributions). This method was used by Wiegell, Tuch, Larsson, and Wedeen (2000).
- Mardia, Kent, and Bibby (1979, page 365) reports a rule of thumb: $K = \sqrt{N/2}$. This is presumably giving a too large number of groups when clustering voxels in functional neuroimaging, — even if the number of resels is used as N .
- Probabilistic assumptions can be made in cluster analysis and if the clusters can be considered a mixture of Gaussians, then generalization estimates can be constructed, either algebraic or test/validation set based:
 - Goutte, Hansen, Liptrot, and Rostrup (2001, Appendix) suggest some algebraic generalization estimates: AIC, BIC, and ICL. Fraley and Raftery (1998) use BIC.
 - Validation set based generalization assessment is performed in (Balslev et al. 2000).

Christiansen (2000) examined both AIC, BIC and cross-validation. Smyth (1996) compared Monte Carlo cross-validation, v -fold cross-validation and BIC with the result generated by the AutoClass program on both simulated as well as non-simulated data. On simulated data v -fold cross validation performed poorly compared to the other methods for identification of the true number of clusters (Smyth 1996, table 1). However, on the non-simulated data which included benchmark data set (iris, diabetes, vowel) it was the best (Smyth 1996, table 2).

Fadili, Ruan, Bloyet, and Mazoyer (2000) describe a method for determining the number of clusters and the fuzziness in fuzzy clustering.

The model selection problem (in cluster analysis the number of clusters) is a never ending issue in statistical modeling. Not too much trust should be put on a single value returned as the “true” number of clusters: The result depends on the metric, e.g., the parametrization of the covariance matrix for Gaussian clusters, see figure 3.4. In the Lyngby toolbox is presently implemented the simple technique of (Thorndike 1953). It should be noted that Biernacki, Celeux, and Govaert (2000) have argued that model selection for density estimation with mixture models and the assessment of the number of clusters are separate issues. Based on their own simulated data set they found that optimal numbers for density estimates are higher than the optimal number of clusters.

Method	Variable	Reference
Dynamic	fMRI time-series	Baune et al. (1999)
Fuzzy	Normalized fMRI data	Scarth et al., (1995, 2000).
K-means	(Short) fMRI time-series	Ding et al. (1994)
K-means with t-test	fMRI time-series	Ding et al. (1996)
K-means/hierarchical	Cross-correlation with paradigm	Toft et al. (1997), Goutte et al. (1999b)
Divisive hierarchical with K-means	(Short) fMRI time-series	Filzmoser et al. (1999)
K-means	Features of the fMRI time-series	Goutte et al., (1998, 2001)
K-means with spatial regularization	MR diffusion images.	Wiegell et al. (2000)
K-means	11 PET images from learning study	Balslev et al. (2000)
K-means/Hierarchical	4 PET images: CBF, CMRO ₂	Toyama et al. (1999, 2001)

Table 3.13: Clustering in functional neuroimaging. See also the “Clustering fMRI time-series” bibliography by Cyril Goutte, <http://eivind.imm.dtu.dk/staff/goutte/fmriclusterefs.html>.

3.8.2 Spatial prior in clustering

It is possible to include position weights so that voxel close to each other would have a tendency to get in the same cluster (Almeida and Ledberg 2001). The resulting cluster volumes will be more spatially homogeneous. This can be implemented by simply adding three extra rows where the location is specified, e.g., the Talairach coordinates in some unit. The scaling between the position information and the other values are important, — in the Lyngby toolbox (Hansen et al. 1999b) the position weight unfortunately have to be set a prior. Another method for obtaining spatial homogeneous clusters is by including the neighboring voxels as features in the vector to be used for clustering (Somorjai, Vivanco, Pizzi, and Jarmasz 2001), — a method that has also been used for canonical correlation analysis (Friman, Cedefamn, Lundberg, Borga, and Knutsson 2001b). The clustering is performed in a larger space with an unfortunate increase in computational demand.

Spatial smoothing of the volumes will produce a similar effect as using spatial priors on the level of clustering, e.g., Balslev et al. (2000) uses spatial smoothing with no spatial priors in the clustering and the resulting clusters are spatially very homogeneous.

The approach of using spatial priors with the K-means algorithm has been used connection with automatic identifications of thalamic nuclei from diffusion tensor images (Wiegell et al. 2000).

3.8.3 Cluster analysis in functional neuroimaging

An early application of cluster analysis for dynamical PET is Ashburner et al. (1996). Others are (Myers et al. 1999; Ho et al. 1997; Feng et al. 1999) for dynamic PET and the two-stage study of Gunn et al. (1999) with fMRI, PET and simulated dynamic PET that first uses cluster analysis to segment the images and after that used *t*-statistics for the segmented regions. Balslev et al. (2000) also use a two-stage analysis with K-means clustering where the cluster centers were found first and then this cluster center were used as a covariate in an *t*-test. Baumgartner et al. (1999) did a similar study using fuzzy clustering analysis. It is important that the explorative and the confirmative step are separated and independent, as the *t*-test will not be valid when fitting has occurred on the same data set. In (Balslev et al. 2000) the data were split across 18 subjects with 9 subjects in the explorative clustering K-means stage, and the rest in the confirmatory *t*-test stage.

Toyama et al. (1999, Toyama et al. (2001) used cluster analysis on four sets of PET images: resting CBF images, hyperventilatory response, acetazolamide response (all with a H₂¹⁵O tracer) and CMRO₂ (with a ¹⁵O gas tracer).

Cluster analysis for fMRI has been applied much more than for PET, — with most of the articles describing the methodology rather than new neuroscientific findings. An overview of the references is available in table 3.13.

Rather than cluster voxels Longstreth Jr et al. (2001) clustered features from structural MRI scans. These were obtained from a population of 5888 elderly and 5 clusters separated the participants in a normal group and in four other groups based on clinical manifestations (e.g., complexity of infarct).

3.9 Principal component analysis and singular value decomposition

Principal component analysis (PCA) (Pearson 1901) is one of the most used unsupervised multivariate analysis methods. It is also called Hotelling transform after Hotelling (1933), Karhunen-Loewe transformation (Karhunen 1947; Loève 1963) and (if the data is frequency data) correspondence analysis. It can be used to summarize the data (by reducing its dimensionality), as a preprocessing step for regression or classification, or for probability density modeling. Thorough

description of PCA are available in Mardia et al. (1979, chapter 8) and Jackson (1991). It can be approached by several definitions (Tipping and Bishop 1997):

- PCA is an orthogonal linear projection that minimizes the squared reconstruction error. According to Tipping and Bishop (1997) this is the property that Pearson (1901) focuses on.
- PCA is a standardized linear projection that maximizes the variance in the projected space (Hotelling 1933), see also Mardia et al. (1979, theorems 8.2.2 and 8.2.3) and theorem B.1.1 on page 145.
- PCA can be obtained from the connectionistic Oja-Sanger rule (Hertz et al. 1991, section 8.3).
- PCA is the maximum likelihood estimator of a (certain) factor analysis model with adjoining probability density (Tipping and Bishop 1997).

The most straight forward method to obtain the principal component of a data matrix is by singular value decomposition (SVD) (Beltrami 1873), also called Eckart-Young-decompositions (Eckardt and Young 1939):

$$\mathbf{ULV}^T = \text{svd}[\mathbf{X}]. \quad (3.21)$$

$\mathbf{U}(N \times R)$ and $\mathbf{V}(P \times R)$ are column orthonormal vectors and called the left and right singular vectors, respectively. $\mathbf{L}(R \times R)$ is a diagonal matrix with the sorted singular values. These vectors correspond to the vectors from the symmetric eigenvalue decomposition or spectral (Jordan) decompositions of the inner and outer product matrix (Mardia et al. 1979, theorem A.6.5)

$$\mathbf{VL}^2\mathbf{V}^T = \mathbf{X}^T\mathbf{X}, \quad (3.22)$$

$$\mathbf{UL}^2\mathbf{U}^T = \mathbf{XX}^T. \quad (3.23)$$

If the data matrix is an indicator matrix then according to Jackson (1991, page 225) the matrix $\mathbf{X}^T\mathbf{X}$ is called a *Burt matrix* after Burt (1950). The transformation from an observation (row) in the data matrix \mathbf{X} to another space ($\mathbf{x} \rightarrow \mathbf{y}$) is called the *principal component transformation*

$$\mathbf{Y} = \mathbf{XV}, \quad (3.24)$$

the column vectors in $\mathbf{Y}(N \times R)$ being called the *principal components* and the columns of \mathbf{V} called the vectors of *principal component loadings*.

There are several ways to “preprocess” (affine transform) the data matrix before performing the SVD. Usually the variable (column) means of the data matrix are extracted

$$\mathbf{ULV}^T = \text{svd}[\mathbf{HX}], \quad (3.25)$$

$\mathbf{H}(N \times N)$ being a centering matrix. The data matrix can also be double-centered by both extracting the row and column means

$$\mathbf{ULV}^T = \text{svd}[\mathbf{H}_N\mathbf{XH}_P], \quad (3.26)$$

which is called M-analysis (Gollob 1968).

If the variables in the data matrix \mathbf{X} have different units they should be scaled, e.g., by their standard deviation

$$\mathbf{ULV}^T = \text{svd}[\mathbf{HXD}^{-1}], \quad (3.27)$$

where $\mathbf{D}(P \times P)$ is a diagonal matrix containing the standard deviations of each variable. Within correspondence analysis, which usually deals with principal component analysis of indicator matrices, there is a special kind of scaling (Mardia et al. 1979, section 8.5) and (Jackson 1991, section 10.8). Diagonal matrices with the row and column sums of the data matrix are constructed

$$\mathbf{D}_N = \text{diag}(\mathbf{1}^T\mathbf{X}) \quad (3.28)$$

$$\mathbf{D}_P = \text{diag}(\mathbf{X}\mathbf{1}), \quad (3.29)$$

which is the applied for the scaling

$$\mathbf{ULV}^T = \text{svd}[\mathbf{D}_N^{-1/2}\mathbf{XD}_P^{-1/2}], \quad (3.30)$$

obtaining what is referred to as *chi-square metric* (Jackson 1991, page 216). In this case the first vector will be proportional to $\mathbf{1}$ and is called the “trivial vector”.

If the data matrix is large the SVD may be too slow or require too much memory. In such a case one can use an iterative version of SVD/PCA, e.g. the Oja-Sanger rule (Hertz et al. 1991, section 8.3). If one of the dimensions N or P

is small symmetric eigenvalue composition on either the inner or outer product covariance (equations 3.22 and 3.23) can be used to obtain the principal components. It is then possible to convert from either \mathbf{U} to \mathbf{V} or the other way

$$\mathbf{U} = \mathbf{X}\mathbf{V}\mathbf{L}^{-1} \quad (3.31)$$

$$\mathbf{V} = \mathbf{X}^T\mathbf{U}\mathbf{L}^{-1}. \quad (3.32)$$

It is possible to test some hypotheses about the principal components (Mardia, Kent, and Bibby 1979, section 8.4). One such test is the isotropy test where it is tested whether the last L eigenvalues are equal. The so-called “Bartlett’s approximation” is an asymptotic χ^2 approximation to the test statistics. This test does not work for singular data. The test is usually said to determine the size of a proper signal subspace for the data: The last equal eigenvalues are assumed to represent noise.

Another method to determine the subspace dimension for singular data was used in (Friston, Frith, Frackowiak, and Turner 1995b) in connection with canonical variate analysis. Here the eigenvectors associated with eigenvalues larger than unity were used after the eigenvalues have been scaled (Friston 2001; Friston, Poline, Holmes, Frith, and Frackowiak 1996b)

$$\tilde{\lambda} = N/(\lambda^T \mathbf{1})\lambda \quad (3.33)$$

with λ as the squared singular values \mathbf{I}^2 . This selects the eigenvectors associated with eigenvalues larger than the mean.

3.9.1 Probabilistic PCA

The decomposition of equation 3.25 is natural if the data is Gaussian distributed $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$: The mean is estimated as $\hat{\boldsymbol{\mu}} = \bar{\mathbf{x}}$ and the covariance as $\hat{\boldsymbol{\Sigma}} = \mathbf{S} = N^{-1}\mathbf{V}\mathbf{L}^2\mathbf{V}$, where \mathbf{V} is computed from equation 3.25.

The model can be augmented by adding a term representing noise to the diagonal of the covariance matrix (Hansen et al. 1997; Hansen et al. 1999a)

$$\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma} + \sigma^2\mathbf{I}) \quad (3.34)$$

This has been called “sensible principal component analysis (SPCA)” (Roweis 1998) and “probabilistic principal component analysis” (Tipping and Bishop 1997; Bishop 1999). By this addition tests can be made on singular data.

3.9.2 Factor analysis

Factor analysis (Mardia, Kent, and Bibby 1979, section 9), (Anderson 1984, chapter 14) or (Conradsen 1984a, section 8.3) is associated with principal component analysis. The “factor model” can be written as²:

$$\mathbf{X} = \mathbf{F}\boldsymbol{\Lambda} + \mathbf{U} + \mathbf{1}\boldsymbol{\mu}^T \quad (3.35)$$

\mathbf{X} is an observed data matrix. \mathbf{F} is a hidden state matrix called the *common factors* or *factor scores*. $\boldsymbol{\Lambda}$ is a matrix of *factor loadings*. \mathbf{U} are the *specific* or *unique* factors. An assumption about uncorrelated common factors $\mathbf{F}^T\mathbf{F}$ is made and the covariance matrix of the unique factors (which in some cases can be considered noise) are diagonal $\mathbf{V}[\mathbf{u}] = \mathbf{\Gamma}$. If this covariance moreover is proportional to the identity matrix $\mathbf{V}[\mathbf{u}] = \sigma_2\mathbf{I}$ then factor analysis corresponds to probabilistic principal component analysis.

The classical MLE for factor analysis is by Jöreskog (1967), see (Mardia, Kent, and Bibby 1979, section 9.4). EM MLE for factor have also been developed (Rubin and Thayer 1982).

3.9.3 Multidimensional scaling

The typical goal of multidimensional scaling (MDS) (as well as self-organizing maps (SOM) Kohonen 1995) is to construct a 2D or 3D space that as good as possible summarize the similarity between a set of data points. The result is often plotted.

MDS will often start from a distance or similarity matrix. In the case where the distance matrix is Euclidean or a distance matrix is constructed from a data matrix by using the square distance the “classical”, “metric” or “principal coordinate analysis (PCoA)” (Gower 1966) MDS solution will be similar to principal components analysis with a row centered data matrix (Mardia, Kent, and Bibby 1979, section 14.3):

$$\mathbf{U}\boldsymbol{\Lambda}\mathbf{V}^T = \text{svd}[\mathbf{H}\mathbf{X}] \quad (3.36)$$

In MDS the vectors in \mathbf{U} is called the principal coordinates.

Similar plots as the MDS plots can be obtained with canonical analysis, see e.g., (Strother et al. 1996a).

²Usually the factor model is written as $\mathbf{x} = \mathbf{A}\mathbf{f} + \mathbf{u} + \boldsymbol{\mu}$, see e.g. (Mardia, Kent, and Bibby 1979; Anderson 1984), but to make the matrices \mathbf{X} and \mathbf{F} of size observation times variables the notation is changed.

3.9.4 SVD/PCA in functional neuroimaging

Application of SVD/PCA in functional brain mapping on voxel-based data are (Friston et al. 1993; Strother et al. 1995a; Strother et al. 1995b; Lautrup et al. 1995; Andersen et al. 1999). The columns vectors of \mathbf{V} are often called *eigenimages*. The scans are often chronologically ordered, and it is then suggested to call the column vectors of \mathbf{U} for *eigensequences*, — this is the term used in the Lyngby toolbox. The “scaled subprofile analysis” (SSM) resembles PCA analysis (Moeller, Strother, Sidtis, and Rottenberg 1987; Moeller and Strother 1991).

SVD has also been applied with EEG (Friedrich, Fuchs, and Haken 1991), MEG (Fuchs, Kelso, and Haken 1992), Multi-electrode recordings (MayerKress, Barcys, and Freeman 1991) and Stetter et al. (2000) describe principal component analysis of OIS images.

Singular value decomposition can be used to obtain the principal components, but as the number of scans N are relatively small the solution can be faster computed with symmetric eigenvalue decomposition of the outer product matrix (R-analysis), see, e.g., (Friston, Frith, Frackowiak, and Turner 1995b; Weaver 1995; Nielsen 1996a) and equation 3.23.

Metric MDS has been used in functional neuroimaging (Friston, Frith, Fletcher, Liddle, and Frackowiak 1996a), (Friston 1997b, section III.D), though this just correspond to singular value decomposition.

3.10 Non-negative matrix factorization

Non-negative matrix factorization (NMF) (Lee and Seung 1999; Lee and Seung 2001) considers the factorization of a positive data matrix $\mathbf{X}(N \times P)$ into two matrices

$$\mathbf{WH} = \text{nmf}(\mathbf{X}), \quad (3.37)$$

where $\mathbf{W}(N \times K)$ and $\mathbf{H}(K \times P)$ are also positive matrices and K is the size of the hidden space. The cost function is symmetric in P and N , thus if the data matrix \mathbf{X} is transposed the estimated matrices \mathbf{W} and \mathbf{H} will just be interchanged, — the same property SVD has. Lee and Seung (2001) suggest two cost functions — an “Euclidean” (squared Frobenius norm) and a divergence-type³

$$E_{\text{“eucl”}} = \|\mathbf{X} - \mathbf{Y}\|_F^2 = \text{trace} \left([\mathbf{X} - \mathbf{Y}]^T [\mathbf{X} - \mathbf{Y}] \right) = \sum_{np} (x_{np} - y_{np})^2 \quad (3.38)$$

$$E_{\text{div}} = D(\mathbf{X}||\mathbf{Y}) = \sum_{np} \left[x_{np} \log \frac{x_{np}}{y_{np}} - x_{np} + y_{np} \right], \quad (3.39)$$

where the matrix $\mathbf{Y} = \mathbf{WH}$. Lee and Seung (2001) also describes an update used for an iterative estimation of the parameters in \mathbf{W} and \mathbf{H} . For the Euclidean cost function this is

$$\mathbf{H}_{kp} \leftarrow \mathbf{H}_{kp} \frac{(\mathbf{W}^T \mathbf{X})_{kp}}{(\mathbf{W}^T \mathbf{WH})_{kp}} \quad (3.40)$$

$$\mathbf{W}_{nk} \leftarrow \mathbf{W}_{nk} \frac{(\mathbf{XH}^T)_{nk}}{(\mathbf{WHH}^T)_{nk}}. \quad (3.41)$$

All the matrix multiplications generate small-sized matrices with one or both dimensions being K , thus the algorithm can work directly on a large data matrix. For the divergence cost function the update is

$$\mathbf{H}_{kp} \leftarrow \mathbf{H}_{kp} \frac{\sum_n \mathbf{W}_{nk} \mathbf{X}_{np} / (\mathbf{WH})_{np}}{(\mathbf{W}^T \mathbf{1}_N^T \mathbf{1}_P)_{kp}} \quad (3.42)$$

$$\mathbf{W}_{nk} \leftarrow \mathbf{W}_{nk} \frac{\sum_p \mathbf{H}_{kp} \mathbf{X}_{np} / (\mathbf{WH})_{np}}{(\mathbf{1}_N \mathbf{1}_P^T \mathbf{H}^T)_{nk}}. \quad (3.43)$$

The update is a gradient type with each parameter scaled individually as in the pseudo-Gauss-Newton. The matrices \mathbf{W} and \mathbf{H} can be initialized as random positive matrices, and both matrices end up being sparse by this algorithm.

Both PET and fMRI raw data are only positive and a non-negative prior for would therefore be appropriate. This new algorithm has not been applied in functional neuroimaging analysis. Hard threshold clusters are obtained by subjecting the matrices to a winner-take-all over the K dimensions: $\mathbf{H}_{\text{wta}} = \text{wta}_K(\mathbf{H})$, and section 5.8.1 presents an application for clustering in author cocitation.

³An other appear in (Lee and Seung 1999) which is just the negative of the divergence apart from a constant: $E = \sum_{np} [\mathbf{X}_{np} \log(\mathbf{WH})_{np} - (\mathbf{WH})_{np}]$.

3.11 Independent Component Analysis

Independent component analysis (ICA) seeks to recover independent hidden source signals from observed multivariate signals that contain the hidden signals mixed linearly (though sometimes nonlinear mixing is considered). With $\mathbf{X}(N \times P)$ as the observed data matrix, $\mathbf{Z}(N \times K)$ as the unobserved source matrix and $\mathbf{A}(K \times P)$ as the unobserved mixing matrix the classic equation is⁴

$$\mathbf{Z}\mathbf{A} = \text{ica}(\mathbf{X}), \quad (3.44)$$

with the fundamental assumption that the K variables in \mathbf{Z} are independent. If the parameters in \mathbf{Z} and \mathbf{A} should be estimable it is necessary to make the extra assumption that the sources \mathbf{z} should be non-Gaussian or that they have different auto-correlation functions. In singular value decomposition the transposition of the data matrix does not matter for the result, — other than the left and right singular vectors are interchanged. However, for ICA it does matter. There are several extension to the “classical ICA” algorithm, as it makes several assumptions:

- There are no autocorrelation in each source signal, i.e., the source signals are white, and the observations are without time-delay or filtering. Other ICA algorithm exists that take advantage of the auto-correlation functions, namely the “Molgedey-Schuster” ICA (Molgedey and Schuster 1994) that does not require that the sources are non-Gaussian. Other algorithm to handle colored sources are dynamic component analysis (DCA) (Attias and Schreiner 1998) and (Amari 2000).
- The dimension of the observation space P is the same as the dimension of the hidden source space, i.e., the number of columns in \mathbf{X} and \mathbf{Z} are equal: $P = K$. The cases where they are not equal has been called over- and under-complete ICA. There exist several methods to deal with these situations, e.g., (Hansen and Larsen 1998; Amari 1999).
- The sources are heavy-tailed (super-Gaussian, fat-tailed, leptokurtic) distributed, e.g., Laplace distributed, as opposed to sub-Gaussian, e.g., uniform distributed. Most signal seems to exhibit heavy-tailed behavior, however there are signals that are sub-Gaussian, e.g. Lee (1998, section 7.3.1) notes that line noise from an EEG measurement has a sub-Gaussian distribution and “could not be clearly isolated by the original infomax algorithm.”
- The observations \mathbf{X} are without noise.

Other complications are nonlinear mixing and non-stationary sources and mixing matrices, furthermore, some algorithms require that the data is preprocess before the actual ICA, e.g., the data is centered and prewhitened.

3.11.1 Molgedey-Schuster ICA

“Molgedey-Schuster” ICA (Molgedey and Schuster 1994) (MS-ICA) finds the independent components when a shifted version of the data exists and there is autocorrelation in the data. It is related to the max/min autocorrelation factor transformation (MAF) (Switzer and Green 1984). The shift is usually either in time or space. This MS-ICA algorithm is very convenient since the mixing matrix \mathbf{A} can be identified directly with a (generalized) eigenvalue decomposition. With the τ time/pixel/voxel step shifted data matrix termed \mathbf{X}_τ the mixing matrix \mathbf{A} is identified as:

$$\mathbf{W} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{X}_\tau, \quad (3.45)$$

$$\mathbf{A}^{-1} \mathbf{L} \mathbf{A} = \text{leig}[\mathbf{W}], \quad (3.46)$$

with the leig as the left eigenvalue decomposition of \mathbf{W} . Section B.5 shows a derivation of this. As the result can become complex it has been suggested (Stetter et al. 2000; Hansen et al. 2000a) to use a symmetric covariance estimate of the last term

$$\mathbf{W} = 1/2(\mathbf{X}^\top \mathbf{X})^{-1} [\mathbf{X}^\top \mathbf{X}_\tau + \mathbf{X}_\tau^\top \mathbf{X}] \quad (3.47)$$

The shift parameter τ represents a parameter that should be set appropriately. A method sets τ to the shift corresponding to the measure of the largest difference in source \mathbf{Z} autocorrelation coefficient (Thomas Kolenda, personal communication). SVD and MS-ICA can be combined (Hansen and Larsen 1998; Stetter et al. 2000; Petersen 2000; Hansen et al. 2001a).

When Molgedey-Schuster ICA is performed with a preliminary SVD the source $\hat{\mathbf{Z}}$ and mixing $\hat{\mathbf{A}}$ matrices can be identified with the following operations

$$\mathbf{U}\mathbf{L}\mathbf{V}^\top = \text{svd}_K(\mathbf{X}) \quad (3.48)$$

$$\tilde{\mathbf{A}}^{-1} \tilde{\mathbf{\Gamma}} \tilde{\mathbf{A}} = \text{leig}(\mathbf{U}^\top \mathbf{U}_\tau + \mathbf{U}_\tau^\top \mathbf{U}) \quad (3.49)$$

$$\hat{\mathbf{A}} = \tilde{\mathbf{A}} \mathbf{L} \mathbf{V}^\top \quad (3.50)$$

$$\hat{\mathbf{Z}} = \mathbf{U} \hat{\mathbf{A}}^{-1}, \quad (3.51)$$

⁴The notation here is a transposed version of the usual: $\mathbf{X}(P \times N) = \mathbf{A}(P \times K)\mathbf{S}(K \times N)$.

where \mathbf{U}_τ is the matrix containing τ -shift eigensequences (if the shift is performed in the time domain), see appendix B.5 and (Hansen and Larsen 1998).

MS-ICA seems to be relatively robust, e.g., Fabricius, Kidmose, and Hansen (2001) found it robust in connection with probability distributions with diverging second moment.

3.11.2 ICA in functional neuroimaging

ICA has been applied both to find the relevant (paradigm) signal as well as confounding signals. The two different versions of ICA that depend on how the data matrix is transposed have been called spatial and temporal ICA:

- For “spatial ICA” (sICA) the images are (assumed to be) independent and there is no constraint on the temporal patterns. This was first applied on fMRI by (McKeown et al. 1998a; McKeown et al. 1998b)
- For “temporal ICA” (tICA) the temporal sequences are (assumed to be) independent and there is no constraint on the images.

Stone, Porrill, Büchel, and Friston (1999a) find that of PCA, sICA and tICA only tICA was able to separate the relevant (they call it “consistent task-related”) temporal activity (i.e. the paradigm) in a single component. For PCA and sICA the relevant signal was mixed up in several components. However, McKeown, Jung, Makeig, Brown, Kindermann, Lee, and Sejnowski (1998a) find that sICA does indeed also separate the relevant component in a single independent component, and they also find the relevant signal is distributed in several components when using PCA.

Petersen (2000) investigates three algorithms Bell-Sejnowski, Molgedey-Schuster and DCA and find that the Molgedey-Schuster algorithm strongly separates the paradigm signal from the other signals and is orders of magnitude faster than Bell-Sejnowski and DCA. Other applications of ICA in functional neuroimaging are (Kolenda 1998) which uses the ICA algorithm of (Bell and Sejnowski 1995) on both PET and fMRI. (Stone, Porrill, Porter, and Hunkin 2000; Stone, Porrill, Büchel, and Friston 1999b) describe a combination of sICA and tICA called spatiotemporal ICA (stICA).

Schiessl et al. (1999), Schiessl et al. (2000) and Stetter et al. (2000) use ICA with preliminary PCA on optical imaging data.

A short review of independent component analysis in functional neuroimaging is available in (Petersson, Nichols, Poline, and Holmes 1999a).

3.11.3 Examples on Molgedey-Schuster ICA

Figure 3.5 shows a Molgedey-Schuster ICA performed on a simple event-related fMRI toy problem: With a “scanning” frequency of $TR = 1s$ 50 events are distributed randomly in a period of 1000s, i.e., the events can be regarded as generated from a Poisson process with Laplacian distributed interstimulus intervals (ISI). The first 9 events are visible as stems in the upper left plot of figure 3.5 which displays the first 150s of the period. Each event is convolved with the Glover/SPM99 hemodynamic response function (see figure 2.5) generating the curve in the upper left plot. Two other signals are present: A sinusoid with approximately 0.9Hz (pseudo-cardiac and heavily aliased) and another with approximately 0.2Hz (pseudo-respiratory), both with a stochastic low-frequency drift in the fundamental frequency. These are the second and third row plots in the left column. From these sources $\mathbf{Z}(N \times K) = \mathbf{Z}(1000 \times 3)$ a small brain scanning data set $\mathbf{X}(N \times P) = \mathbf{X}(1000 \times 30)$ is constructed $\mathbf{X} = \mathbf{Z}\mathbf{A} + \mathbf{U}$ by a random mixing matrix $\mathbf{A}(K \times P)$ and the disturbance term \mathbf{U} containing Gaussian noise. Molgedey-Schuster ICA is performed $\hat{\mathbf{Z}}\hat{\mathbf{A}} = \text{msica}_K(\mathbf{X})$ with a shift of $\tau = 2$ and the recovered sources $\hat{\mathbf{Z}}$ shown in the right column of figure 3.5: Apart from permutation, sign changes and magnitude the sources are well recovered. Indeed in this simple example the sources are also separable with ordinary singular value decomposition.

When the stimuli are generated from a Poisson process its auto-correlation function will be a Dirac delta function in zero. The separation of the stimulus-related component in the example relies on the autocorrelation of hemodynamic response and if confounds in the scans are convolved with the hemodynamic response function the separation of the sources will immediately deteriorate.

Figure 3.6 displays the result of a Molgedey-Schuster ICA on a real fMRI data acquired by Egill Rostrup, Hvidovre Hospital with a 1.5 Tesla scanner: A single subject block fMRI study with an inverting full-field 8 Hz-checkerboard stimuli that was also used in (Goutte, Toft, Rostrup, Nielsen, and Hansen 1999b; Goutte, Hansen, Liptrot, and Rostrup 2001). $TR = 3Hz$ and 10 runs of 150 scans were obtained with a paradigm of rest (20 sec, 60 scans) — activation (10 sec, 30 scans) — rest (20 sec, 60 scans). The data covers a single slice from two different sessions with the same subject. Cyril Goutte performed linear trends removal from each voxel in each individual run and F-masking was applied to discard voxels that only contained noise, resulting in 3891 and 4372 voxels for each session. 5 of the 10 runs were used for ICA and the first and last 25 from each run were left out leaving 605 scans in each session. The corresponding images from each session were concatenated into one long vector: $\mathbf{X}(605 \times 8263) = [\mathbf{X}_1(605 \times 3891), \mathbf{X}_2(605 \times 4372)]$. The number of components were chosen to $K = 8$ while the time shift was set to $\tau = 1$.

The paradigm is clearly separated in one component (the lower row): From the time domain plot of the source the periodic paradigm is easily seen. The mixing matrix for this component displays a signal in the visual cortex (the bottom

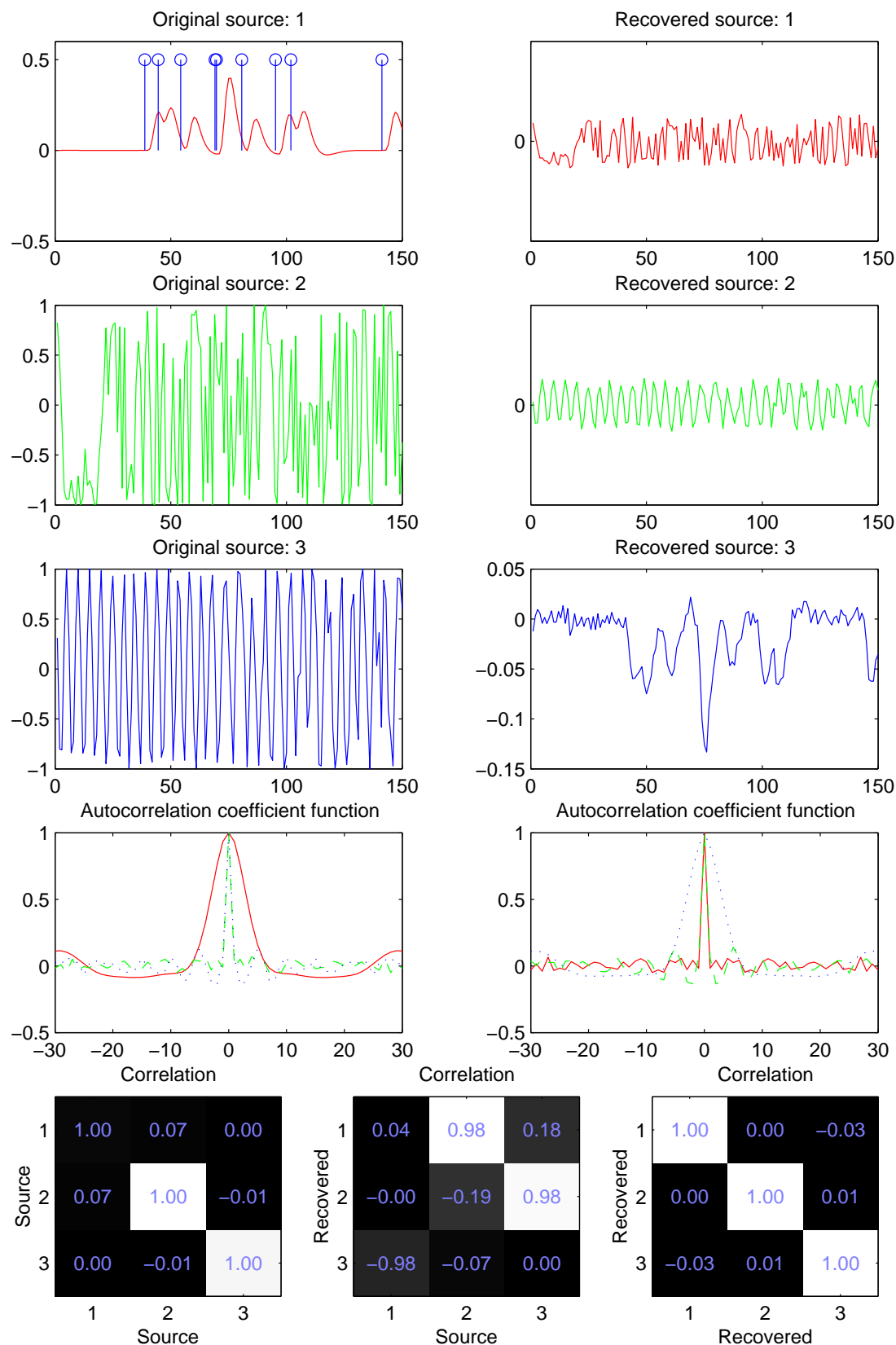


Figure 3.5: Demonstration of Molgedey-Schuster ICA on a toy problem. Subplots of the left column correspond to original source signals and right columns to recovered sources. Note that the ordering between the original and the recovered sources are permuted, since the ICA algorithm is not aware of the original ordering: The “paradigm” signal source appears in the first row while it is recovered in the third (with inverted sign). The “cardiac” original signal source is shown in the second row and covered as the first row, while the “respiratory” original signal appears as the third row and recovered in the second row.

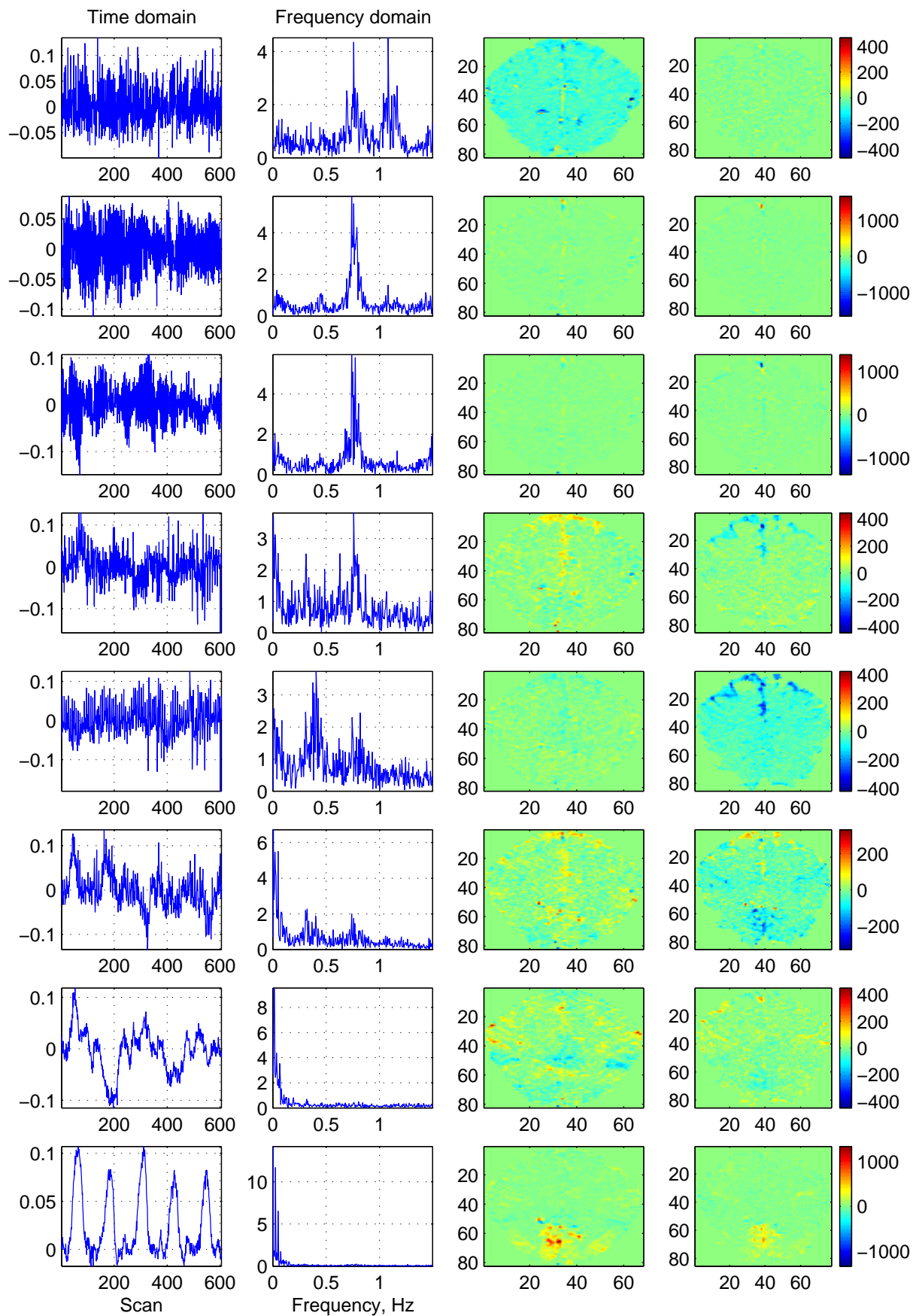


Figure 3.6: Demonstration of Molgedey-Schuster ICA on real fMRI data. Each row corresponds to a single independent component. The first and second column display the sources in the time domain and in the frequency domain, respectively. The third and fourth column display the mixing matrix split on the two sessions. Note that the aspect ratios are not natural.

of the image) that are uneven distributed between the first and second session — the first session showing a larger signal both in the sense of magnitude and spatial extent. This might very well constitute a difference in the fMRI signal, as also reported by McGonigle, Howseman, Athwal, Friston, Frackowiak, and Holmes (2000).

The components corresponding to the first, second and third row have a strong periodic signal around 0.76 Hz. This correspond to a slow cardiac frequency of approximately 46 beats per minute. The signal is spatial localized in the top of the image — the anterior part of the superior sagittal sinus. This shows that the Molgedey-Schuster ICA is able to identify a confound in line with what others have reported for other types of ICA algorithms, see table 3.10. It should be noted that the K-means cluster analysis performed on the cross-correlation function with the same data (Goutte, Toft, Rostrup, Nielsen, and Hansen 1999b) did not find such a clearly separated component. It should also be noted that Biswal and Ulmer (1999) detected no cardiac-related with their ICA based on 3 T fMRI. They attributed this to a long TR of 600 ms.

The first row component contains a second frequency around 1.08 Hz. It is not know what this constitutes, e.g., it does not fit with a direct harmonic or an aliased harmonic. The components for the fifth and sixth row contain an approximated 0.4 Hz signal and its first harmonic, that could be a signal associated with respiration.

3.12 Probability density estimation

Probability density estimation (PDE) considers the problem of estimating a probability density function (PDF) $p(\mathbf{x})$ from a finite number of samples with some form of model $p(\mathbf{x}|\theta) \approx p(\mathbf{x})$. PDE can be used for outlier detection and in prediction, e.g., classification where class-conditional densities are needed. Some of the many variation of PDE can be grouped in the following way:

- **Parametric methods** where the form of the PDF is restricted to a simple often well-known PDF, e.g., the Gaussian distribution. The advantages with these models are that they (usually) consist of few parameters making it unlikely to overfit, and one is often able to derive hypothesis tests to an analytical form. The disadvantage is that it is not very flexible, e.g., not able to model bimodal PDFs.
- **Finite mixture models** are models that use several simple PDFs (kernels) to model the target PDF. A common choice is a weighted additive mixture of a finite number of Gaussian distributions. The advantages with these models are that they are relatively flexible and that the individual mixture components might lend themselves to interpretation, i.e., one is able to identify clusters in the target PDF. The disadvantages are that it is not possible (or difficult) to derive analytical expression for hypothesis testing, the estimation of the parameters is difficult (the estimation requires iterative updates), and for some cases, e.g., the Gaussian mixture model with no assumption about the covariance, the likelihood can become infinite if a single object absorbs a mixture component exclusively.
- **Kernel methods.** Also called *probabilistic neural networks* (Specht 1990), *Parzen windows* (Parzen 1962) or non-parametric methods (even though they usually consists of very many parameters). In kernel methods each object is convolved with a simple PDF, e.g., a uniform distribution (Rosenblatt 1956) or a Gaussian distribution (Parzen 1962). It can be regarded as a mixture model where the number of mixture components are equal to the number of objects. Unlike the mixture model each “mixture” in a kernel method is not interpretable.
- **Infinite mixture models.** Rasmussen (2000) has developed an PDE model that in principal can have infinite number of mixture components each weighted within a Bayesian scheme. The model is very flexible but the components are not interpretable. Related are the finite mixture model of Richardson and Green (1997) that does not have a fixed number of components.

Apart from the above mentioned there are *orthogonal expansion estimation* which uses basis function such as Fourier series and *projection pursuit density estimation*, which are based on the projection pursuit (Ripley 1996, pages 185-188). Mixture models and kernel density estimation is described in more detail in the following sections.

3.12.1 Mixture models

Finite mixture models estimate a probability density $p(\mathbf{x})$ by several (K) other simpler component densities $p(\mathbf{x}|k)$

$$p(\mathbf{x}) \approx \sum_k^K p(\mathbf{x}|k)P(k) \quad (3.52)$$

where $P(k)$ is a prior for the k 'th component. The component densities are often parametrized as multivariate Gaussian distributions

$$p(\mathbf{x}|k) = |2\pi\mathbf{\Sigma}_k|^{-1/2} \exp \left[-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_k)\mathbf{\Sigma}_k^{-1}(\mathbf{x} - \boldsymbol{\mu}_k)^T \right]. \quad (3.53)$$

The means $\boldsymbol{\mu}_k$ are usually not restricted, but the covariance matrices $\mathbf{\Sigma}_k$ can be parametrized in various ways, see figure 3.4, and they can be assumed to be common $\mathbf{\Sigma}_1 = \dots = \mathbf{\Sigma}_K$ (Hastie and Tibshirani 1996), which avoids the problem

in maximum likelihood estimation where the likelihood goes to infinity if a mixture component collapses to a single data point and overfit dramatically (Day 1969; Bishop 1995a, page 63). The result probability are parametrized as

$$p(\mathbf{X}|\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_K, \boldsymbol{\Sigma}_1, \dots, \boldsymbol{\Sigma}_K, \mathbf{k}) = \prod_n^N p(\mathbf{x}_n|\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_K, \boldsymbol{\Sigma}_1, \dots, \boldsymbol{\Sigma}_K, \mathbf{k}) \quad (3.54)$$

$$= \prod_n^N \left[\sum_k^K p(\mathbf{x}_n|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) P(k) \right] \quad (3.55)$$

where the vector $\mathbf{k} = [P(k=1), \dots, P(k=K)]^\top$ contains the component priors. Differentiating the log-likelihood of this does unfortunately not yield a simple expression that directly can be optimized with one of the methods listed in section 3.2.1, instead the so-called EM-algorithm are often applied (Bishop 1995a, section 2.6.2).

Hansen, Sigurdsson, Kolenda, Nielsen, Kjems, and Larsen (2000d), Nielsen and Hansen (1999) and Christiansen (2000) describe a special EM scheme, where the data is split into two disjoint sets and one set estimates the means $\boldsymbol{\mu}_k$ while the other the covariances $\boldsymbol{\Sigma}_k$. The entire data is used for estimation of the component priors $P(k)$. Each objects \mathbf{x}_n is hard clustered to its assigned component and if there is not enough objects N_k for the estimation of the covariance of the k 'th component this covariance is then estimated as a scaled version of the covariance on the entire data set $\boldsymbol{\Sigma}_0$. The algorithm is as described in (Hansen et al. 2000d) with initialization of the overall parameters $\boldsymbol{\mu}_0$ and $\boldsymbol{\Sigma}_0$ and the individual component parameters $\boldsymbol{\mu}_k$, $\boldsymbol{\Sigma}_k$ and $P(k)$

$$\boldsymbol{\mu}_0 \leftarrow N^{-1} \mathbf{X}^\top \mathbf{1}_N \quad (3.56)$$

$$\boldsymbol{\Sigma}_0 \leftarrow N^{-1} (\mathbf{X} - \mathbf{1} \boldsymbol{\mu}_0^\top)^\top (\mathbf{X} - \mathbf{1} \boldsymbol{\mu}_0^\top) \quad (3.57)$$

$$\boldsymbol{\mu}_k \sim \mathcal{N}(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0). \quad (3.58)$$

$$\boldsymbol{\Sigma}_k \leftarrow \boldsymbol{\Sigma}_0 \quad (3.59)$$

$$P(k) \leftarrow 1/K. \quad (3.60)$$

The following steps are repeated until convergence

$$\mathcal{S}_\mu, \mathcal{S}_\Sigma \leftarrow \text{Split } n = 1..N \text{ randomly into two sets} \quad (3.61)$$

$$\mathcal{K}_k \leftarrow \{n : \max_k p(k|\mathbf{x}_n)\} \quad (3.62)$$

$$\boldsymbol{\mu}_k \leftarrow N_{k,\mu}^{-1} \sum_{n \in \mathcal{K}_k \wedge n \in \mathcal{S}_\mu} \mathbf{x}_n \quad (3.63)$$

$$\boldsymbol{\Sigma}_k \leftarrow \begin{cases} N_{k,\Sigma}^{-1} \sum_{n \in \mathcal{K}_k \wedge n \in \mathcal{S}_\Sigma} (\mathbf{x}_n - \boldsymbol{\mu}_k)(\mathbf{x}_n - \boldsymbol{\mu}_k)^\top & \text{if } N_k > P + 1 \\ \boldsymbol{\Sigma}_0 & \text{if } N_k \leq P + 1 \end{cases} \quad (3.64)$$

$$P(k) \leftarrow N_k/N \quad (3.65)$$

where \mathcal{K}_k is the N_k sized set of assigned objects to the k 'th component. The algorithm is iterated over a number of different K s, e.g., $K = 1..N/2$ and the best model is selected via the AIC criteria. Unless two object are the same the algorithm will avoid the infinite likelihood problem and has therefore been termed the *generalizable Gaussian mixture* (GGM) model. Classification can be made with this model if it is applied separately on each class, see section 5.9.2 with figure 5.17(a) and 5.18 for an application.

3.12.2 Kernel methods

Kernel methods for density estimation place a kernel around each of N P -dimensional data points \mathbf{x}_n in the N -sized training set, i.e., convolve each point with a kernel (Ripley 1993, section 6.1), (Bishop 1995a, section 2.5.3), (Duda, Hart, and Stork 2000, section 4.3). The kernel often has the form of a simple parametric probability density function $p(\mathbf{x}|n)$:

$$p(\mathbf{x}) \approx \sum_n^N p(\mathbf{x}|n) P(n). \quad (3.66)$$

Often each of the N components are given equal weight $P(n) = 1/N$, and frequently the kernel is a Gaussian probability density function with a homogeneous isotropic covariance $\boldsymbol{\Sigma}_n = \sigma^2 \mathbf{I}$

$$p(\mathbf{x}|n) = p(\mathbf{x}|\mathbf{x}_n, \sigma^2) = (2\pi\sigma^2)^{-P/2} \exp\left(-\frac{1}{2\sigma^2}(\mathbf{x} - \mathbf{x}_n)^\top (\mathbf{x} - \mathbf{x}_n)\right) \quad (3.67)$$

where σ^2 can be referred to as the kernel width. If the kernel width is not set from a priori knowledge it might be estimated from the data with the methods mentioned in section 3.3, e.g., via cross-validation (Silverman 1986). The negative log-probability leave-one-out (LOO) cross-validation cost function is

$$E(\mathbf{X}, \sigma^2) = - \sum_n^N \log p_{-n}(\mathbf{x}_n | \mathbf{X}_{-n}, \sigma^2) \quad (3.68)$$

$$= - \sum_n^N \log \left(\frac{1}{N-1} \sum_{n' \neq n}^N (2\pi\sigma^2)^{-P/2} \exp \left[-\frac{1}{2\sigma^2} (\mathbf{x}_n - \mathbf{x}_{n'})^T (\mathbf{x}_n - \mathbf{x}_{n'}) \right] \right). \quad (3.69)$$

The LOO cross-validation function can be differentiated with respect to σ^2 and optimized with the second order Newton method (Nielsen, Hansen, and Kjems 2001b), see figure 3.7 for a simple example.

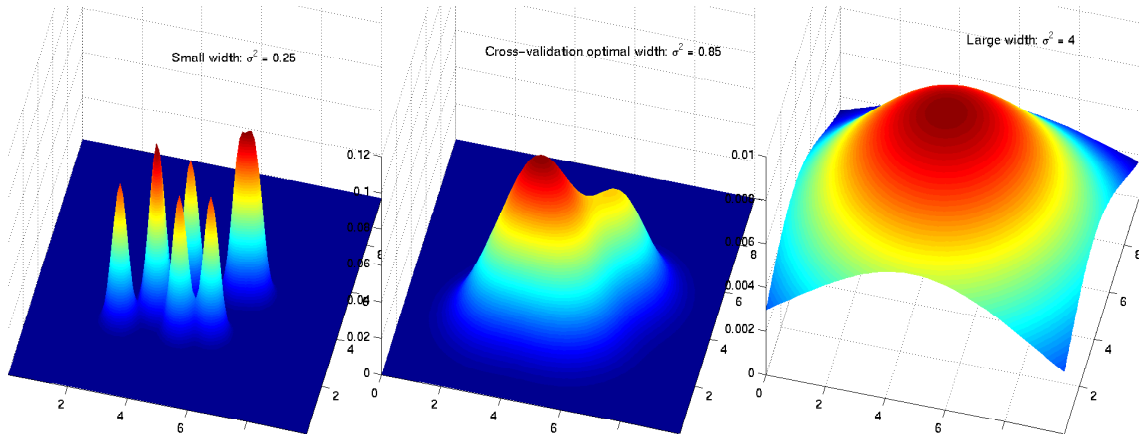


Figure 3.7: Kernel density estimation. Estimation from $N = 8$ data points in $P = 2$ dimensional space with varying kernel width σ^2 . The middle panel is with a LOO cross-validation optimal σ^2 . Compared with this the σ^2 is too small in the left panel and too large in the right panel.

3.12.3 Probability density estimation in functional neuroimaging

Probability density modeling has been used to model the hot spot location from functional neuroimaging results: Fox et al. (1997a), Fox et al. (1997b), Fox et al. (1999) and Fox et al. (2001) used simple parametric models, while Nielsen and Hansen (1999) worked with finite mixture models. Kernel methods have been applied in (Nielsen, Hansen, and Kjems 2000; Nielsen and Hansen 2000b; Nielsen and Hansen 2001b; Nielsen, Hansen, and Kjems 2001b; Turkeltaub, Eden, Jones, and Zeffiro 2001). Hartvig and Jensen (2000) used mixture models in the spatial domain and Everitt and Bullmore (1999) performed mixtures for activated and non-activated components.

3.13 Novelty and outlier detection

Novelty and outlier detection are concerned with abnormal samples. Beckman and Cook (1983) define outliers as either discordant (“any observation that appears surprising or discrepant to the investigator”) or contaminant (“any observation that is not a realization from the target distribution”). “Outlier detection” is often used in the statistical literature while “novelty detection” are mostly used in the connectionistic literature.

Novelty detection can be categorized in two classes:

- **Monitoring of data.** Novel data might be interesting (discordant) or faulty (contaminant) and we would like to detect any of these, e.g., “early detection of faults in rotating mechanical machinery” (Ypma and Duin 1998), epileptic and non-epileptic EEG signals (Roberts and Tarassenko 1994; Roberts 1999), tumors in mammograms (Tarassenko, Hayton, Cerneaz, and Brady 1995), or detection of neuroscientific database errors and other novelties (Nielsen, Hansen, and Kjems 2001b; Nielsen and Hansen 2001b).
- **Monitoring of model performance.** Operating the model in a part of the input space where it has never been trained will usually mean that the prediction will be unreliable (have a high variance) (Bishop 1994). A non-stationary input distribution or a limited number of data samples might be the reason why the model is in a novel part of the input space. Bayesian modeling where the whole predictive distribution is modeled will usually give an indication

of the prediction uncertainty: Here no extra modeling is necessary only an extraction of a value (e.g., the variance) from a distribution. Once novel data has been detected the model might be retrained.

For Gaussian multivariate data there are several techniques to detect outliers (Beckman and Cook 1983), e.g., a Mahalanobis distance. For non-Gaussian multivariate distributions more flexible probability density modeling can be used, e.g., Roberts and Tarassenko (1994) and Roberts (1999) use a Gaussian mixture model. Bishop (1994) uses also a kernel density estimator (Parzen window) with a Gaussian kernel and Martinez (1998) uses a “neural tree”. Other unsupervised methods are self-organizing maps with a goodness of map value (Ypma and Duin 1998) and autoencoder/associative network combined with the correlation coefficient or the Euclidean distance between the input and the output (reconstructed) data (Pomerleau 1993; Worden 1997). Instead of the unsupervised technique it is also possible to use supervised, building a model that classifies normal and abnormal samples. However, it is usually not a good idea: The abnormal samples are expensive to collect and the abnormal distribution might be non-stationary.

Building the model for novelty detection is dependent on whether the data set contains only “normal” samples or both normal and the abnormal samples. In the first case the PDF modeling can be modeled from all data samples. In the second case the abnormal cases should not affect the PDF estimate of the normal class. Roberts (1999) suggests that the PDF model-complexity should be penalized (by MDL). For simple parametric models it should be possible to use robust estimation that is less influenced by outliers. In (Nielsen, Hansen, and Kjems 2001b; Nielsen and Hansen 2001b) a heuristic two-stage strategy was used where a probability density first was established and then the 5% of the most extreme data points were discarded and a new probability density was estimated from the remaining data points. This forms an implicit assumption of 5% outliers in the data set. (Larsen, Andersen, Hintz-Madsen, and Hansen 1998a) describes a method where the outlier probability is estimated based on a validation set.

A problem is to get from the probability density associated with an outlier to the probability. In connection with the simple parametric model one is usually able to derive a p-value. For the novelty detection for kernel methods and mixture models this is not straight forward. Hansen, Sigurðsson, Kolenda, Nielsen, Kjems, and Larsen (2000d) describe a method with sorting of the density values and establishing a p-value from the (countable) fraction of data points below the density value. This p-value can be estimated on the training, validation and test set.

Bishop (1994) used posterior odds as a measure for the novelty so that a data point \mathbf{x} is assign to the novelty class if

$$P(\mathcal{C}_1|\mathbf{x}) < p(\mathcal{C}_2|\mathbf{x}), \quad (3.70)$$

where \mathcal{C}_1 denotes the non-contaminated training set and \mathcal{C}_2 denotes the novelty class. These posteriors are found through Bayes formula

$$P(\mathcal{C}_c|\mathbf{x}) = \frac{p(\mathbf{x}|\mathcal{C}_c) P(\mathcal{C}_c)}{p(\mathbf{x})}. \quad (3.71)$$

Since $p(\mathbf{x})$ is the same for both \mathcal{C}_1 and \mathcal{C}_2 the denominator of equation 3.71 will be eliminated when it is plugged into equation 3.70 and it reduces to

$$p(\mathbf{x}|\mathcal{C}_2) > \frac{p(\mathbf{x}|\mathcal{C}_1) p(\mathcal{C}_1)}{p(\mathcal{C}_2)}. \quad (3.72)$$

In (Bishop 1994) the novelty density $p(\mathbf{x}|\mathcal{C}_2)$ “is taken to be constant over some large region of input space” while we in (Nielsen and Hansen 2001b; Nielsen, Hansen, and Kjems 2001b) implicitly used a uniform “improper” distribution $p(\mathbf{x}|\mathcal{C}_2) \propto 1$, and the novelty becomes inverse proportionally related to the posterior of the non-contaminated data $p(\mathbf{x}|\mathcal{C}_1)$. This allowed us to compare novelty indices across the different data sets.

3.14 General model for unsupervised methods

Some of the unsupervised methods can be seen as a specialization of a general model “analysis of covariance structures” proposed by Jöreskog (1970), see also (Mardia et al. 1979, section 9.9). With a data matrix $\mathbf{X}(N \times P)$ each row being Gaussian distributed the mean and covariance are

$$\mathbf{E}[\mathbf{X}] = \mathbf{A}\mathbf{\Xi}\mathbf{P} \quad (3.73)$$

$$\mathbf{V}[\mathbf{x}] = \mathbf{\Sigma} = \mathbf{B}(\mathbf{\Lambda}\mathbf{\Phi}\mathbf{\Lambda}^\top + \mathbf{\Psi})\mathbf{B}^\top + \mathbf{\Theta}. \quad (3.74)$$

This general model says that the mean and the covariance are structured in some way. $\mathbf{\Xi}(G \times H)$ is a matrix of parameters. The semi-positive definite diagonal matrices $\mathbf{\Psi}(Q \times Q)$ and $\mathbf{\Theta}(P \times P)$ will often represent noise of some sort. The symmetric $\mathbf{\Phi}(R \times R)$ on the other hand often describes the interesting signal and $\mathbf{\Lambda}(Q \times R)$ and $\mathbf{B}(P \times Q)$ are the matrices that rotate the $\mathbf{\Phi}$ -space to the space of \mathbf{X} .

In the original work of Jöreskog the matrices $\mathbf{A}(N \times G)$ and $\mathbf{P}(H \times P)$ were regarded as known. However, if \mathbf{A} consists of parameters then mixture models with common covariance among the mixture components can also be incorporated in the general model, e.g., in the K-means algorithm the parameters of \mathbf{A} will be identifiers for assignment,

Model	\mathbf{B}	$\mathbf{\Lambda}$	$\mathbf{\Phi}$	$\mathbf{\Psi}$	$\mathbf{\Theta}$
PCA	\mathbf{I}	$\mathbf{\Lambda}$	\mathbf{I}	$\mathbf{0}$	$\mathbf{0}$
Probabilistic/Sensible PCA	\mathbf{I}	$\mathbf{\Lambda}$	\mathbf{I}	$\sigma\mathbf{I}$	$\mathbf{0}$
Factor model	\mathbf{I}	$\mathbf{\Lambda}$	\mathbf{I}	$\mathbf{\Psi}$	$\mathbf{0}$
“Probabilistic K-means”	$\mathbf{0}$	$\mathbf{0}$	$\mathbf{0}$	$\mathbf{0}$	$\sigma\mathbf{I}$
Bilinear model	\mathbf{I}	(\mathbf{P}, \mathbf{Q})	\mathbf{I}	$\mathbf{\Upsilon}$	$\mathbf{0}$

Table 3.14: Specialization of Jöreskog’s model “analysis of covariance structures”. For the bilinear model see section 3.18.6.

— an assignment matrix. The number of columns in \mathbf{A} (G) will be equal to the number of clusters (K) and $\mathbf{\Xi}$ will contain coordinates for the cluster centers.

Another effort to unify several of the unsupervised models is Roweis and Ghahramani (1999). The basic model is in a notation known from discrete time linear dynamical systems:

$$\mathbf{z}_{t+1} = \mathbf{D}\mathbf{z}_t + \mathbf{w}_t \quad (3.75)$$

$$\mathbf{x}_t = \mathbf{C}\mathbf{z}_t + \mathbf{v}_t. \quad (3.76)$$

This is directly related to Jöreskog’s model, if the discrete time representation is replaced by the introduction of an intermediate space denoted with \mathbf{y}

$$\mathbf{y} = \mathbf{D}\mathbf{z} + \mathbf{w} \quad (3.77)$$

$$\mathbf{x} = \mathbf{C}\mathbf{y} + \mathbf{v}. \quad (3.78)$$

Ignoring any mean of the variables, provided that there is no dependence between noise and data (e.g., $\mathbf{C}[\mathbf{C}\mathbf{y}, \mathbf{v}] = \mathbf{0}$), and applying ordinary covariance matrix manipulation, see e.g., (Mardia et al. 1979, equations 2.2.11 and 2.2.14), the covariance of \mathbf{x} can be identified as

$$\mathbf{V}[\mathbf{x}] = \mathbf{V}[\mathbf{C}\mathbf{y} + \mathbf{v}] \quad (3.79)$$

$$= \mathbf{V}[\mathbf{C}\mathbf{y}] + \mathbf{V}[\mathbf{v}] + \mathbf{C}[\mathbf{C}\mathbf{y}, \mathbf{v}] + \mathbf{C}[\mathbf{v}, \mathbf{C}\mathbf{y}] \quad (3.80)$$

$$= \mathbf{C}\mathbf{V}[\mathbf{y}]\mathbf{C}^T + \mathbf{V}[\mathbf{v}]. \quad (3.81)$$

Further applying the same transformation between \mathbf{y} and \mathbf{z} the result is

$$\mathbf{V}[\mathbf{x}] = \mathbf{C} [\mathbf{D}\mathbf{V}[\mathbf{z}]\mathbf{D}^T + \mathbf{V}[\mathbf{w}]] \mathbf{C}^T + \mathbf{V}[\mathbf{v}]. \quad (3.82)$$

By comparing equation 3.74 and 3.82 the following can be written

$$\mathbf{V}[\mathbf{v}] = \mathbf{\Theta} \quad (3.83)$$

$$\mathbf{V}[\mathbf{w}] = \mathbf{\Psi} \quad (3.84)$$

$$\mathbf{V}[\mathbf{z}] = \mathbf{\Phi} \quad (3.85)$$

$$\mathbf{C} = \mathbf{B} \quad (3.86)$$

$$\mathbf{D} = \mathbf{\Lambda} \quad (3.87)$$

$$(3.88)$$

3.15 Supervised modeling

Supervised models consider two sets of data. One set of data is considered the independent data and is regressed on the dependent data set — the target.

Attributed as supervised modeling are classification and linear as well as nonlinear regression, such as (linear) multivariate regression analysis and multilayer artificial neural networks.

3.16 Linear modeling

3.16.1 Regression

One of the most used linear models is the *multivariate regression model* (Mardia, Kent, and Bibby 1979, chapter 6)

$$\mathbf{Y} = \mathbf{X}\mathbf{B} + \mathbf{U} \quad (3.89)$$

Name	Form	Primary references
Poisson		Friston et al. 1994a
Gaussian	$\mathcal{N}_{\mu, \sigma^2} = \mathcal{N}_{6, 3^2}$	Rajapakse, Kruggel, Maisog, and von Cramon 1998, Hartvig 1999, equation 6
Gamma density	$\beta\gamma(t)$	Lange and Zeger 1997
Delayed gamma	$\beta\gamma(t - \tau)$	Boynton et al. 1996, Dale and Buckner 1997, Ollinger et al. 2001a
Gamma mixture (+derivatives)	$\beta [\gamma_{6,1}(t) + 1/6\gamma_{16,1}(t)]$	Glover 1999, Friston et al. 1998a
Gamma mixture	$\beta_1\gamma(t) + \beta_2\gamma(t - 1) + \dots + \beta_5\gamma(t - 5s)$	Konishi et al. 2001
Fourier series		Friston et al. 1995b, Bullmore et al. 1996a Josephs et al. 1997
FIR	$\beta_1\delta(n) + \beta_2\delta(n - 1) + \beta_K\delta(n - K)$	Nielsen et al. 1997, Goutte et al. 2000, equation 1, Ollinger et al. 2001a, Ollinger et al. 2001b
Volterra series (Gamma mixture)		Friston et al. 2000b
Artificial neural network	$\tanh(\mathbf{x}^T \mathbf{V})\mathbf{w}$	Nielsen et al. 1999

Table 3.15: Hemodynamic response models: linear and nonlinear. γ is the gamma density function.

$\mathbf{Y}(N \times P)$ is an observed data matrix with P variables and N objects and $\mathbf{X}(N \times Q)$ is a known matrix. $\mathbf{B}(P \times Q)$ is a parameter matrix and $\mathbf{U}(N \times P)$ is a matrix of random disturbances or *noise*. If \mathbf{X} is a design matrix then the model is called the *general linear model* (GLM). In functional neuroimaging \mathbf{Y} is often the brain scanning data in a voxel representation. If there is only one variable $P = 1$ then the model is the *multiple regression model*:

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{u}. \quad (3.90)$$

When $P = 1$ it corresponds to modeling each voxel separately. When all voxels are iteratively modeled this way it is often referred to a *mass-univariate* modeling.

The parameters in multiple regression model can be estimated as the maximum likelihood (ML) — with an assumption of independently Gaussian distributed noise — or ordinary least square (OLS)

$$\hat{\mathbf{b}}_{\text{OLS}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}. \quad (3.91)$$

The multiple regression model incorporates numerous subsumed models, e.g., ordinary regression, ANOVA (analysis of variance), ANCOVA (analysis of covariance) and some cases of “ t -tests” and “ F -tests”. Also the finite impulse response (FIR) filter — known from digital signal processing, see, e.g., (Mitra 2001, section 6.3) — is a variation. The type of multiple regression model depends on how the (design) matrix \mathbf{X} is constructed: When group structures are investigated the design matrix \mathbf{X} contains indicators (ones and zeros). In analysis of fMRI basis functions for the hemodynamic response function are often represented in the columns of the design matrix, table 3.15 shows some of the basis functions that have been used.

If there are parameters in the basis functions that need to be fitted then it is not possible to estimate nonlinear parameters with equation 3.91. An example is the gamma density models: while the amplitude parameter in this model can be estimated directly, the nonlinear form and shape parameters cannot be identified direct and should undergo nonlinear optimization. The infinite impulse response (IIR) filter — another model associated with digital signal processing — can neither be incorporated in the multiple regression model.

In the presence of noise, if there are many parameters Q compared to the number of objects N , and/or some of the variables in \mathbf{X} are highly correlated the estimation in the multiple regression model (as in any other model) will be unstable. This is, e.g., the case in connection with FIR filters applied on fast sampled fMRI data. With a TR = 333ms and a filter length corresponding to 30 seconds 100 filter coefficients are required, and with the usual size of the data set the variance of the parameter estimate is high. In this connection regularization can help, e.g., ridge regression where the parameters are identified as (Hoerl and Kennard 1970b; Hoerl and Kennard 1970a)

$$\hat{\mathbf{b}}_{\text{RR}} = [\mathbf{X}^T \mathbf{X} + \lambda \mathbf{I}]^{-1} \mathbf{X}^T \mathbf{y}. \quad (3.92)$$

This can be regarded as a MAP estimate of \mathbf{b} where the prior distribution of \mathbf{b} is a an isotropic and homogeneous Gaussian. The estimate will be *shrunk* toward zero.

Another common regularization method is principal component regression (PCR) (Massy 1965), for a neuroimaging application see, e.g., (Østergaard, Weiskoff, Chesler, Gyldensted, and Rosen 1996). Here the estimation is restricted to a linear subspace obtained by a singular value decomposition of the design matrix (Burnham, Viveros, and MacGregor 1996, equation 6)

$$\hat{\mathbf{b}}_{\text{PCR}} = \mathbf{V}\mathbf{L}^{-2}\mathbf{V}^T\mathbf{X}^T\mathbf{y} \quad (3.93)$$

$$\mathbf{U}\mathbf{L}\mathbf{V}^T = \text{svd}_K[\mathbf{X}], \quad (3.94)$$

where svd_K is a K -truncated singular value decomposition.

A special regularization is proposed in (Goutte, Nielsen, and Hansen 2000; Nielsen, Goutte, and Hansen 2001a) in connection with the so-called “smooth FIR” model. Here smooth kernels (i.e., hemodynamic response function models) are produced by having a Gaussian prior distribution for the parameters

$$p(\mathbf{b}) \propto \exp[-\mathbf{b}\Sigma_{\mathbf{bb}}^{-1}\mathbf{b}/2], \quad (3.95)$$

where the individual parameters b_i in \mathbf{b} next to each other is more correlated, the (i, j) -element set as

$$\Sigma_{\mathbf{bb};i,j} = \nu \exp\left(-\frac{1}{2l}(i-j)^2\right) \quad (3.96)$$

where ν controls the amount of regularization while l is a characteristic length of the smoothness. The MAP estimate of the parameters are then found to be

$$\hat{\mathbf{b}}_{\text{SFIR}} = [\mathbf{X}^T\mathbf{X} + \sigma^2\Sigma_{\mathbf{bb}}^{-1}]^{-1}\mathbf{X}^T\mathbf{y} \quad (3.97)$$

where σ^2 are the noise variance of \mathbf{u} : $\mathbf{u} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$. With $l \rightarrow 0$ the ridge regression solution will be approached.

A similar method is described by Marrelec and Benali (2001) though here the regularization matrix \mathbf{R} seems to be “truncated” to a band-diagonal matrix. The confounds are modeled simultaneously by a polynomial. The Gaussian prior regularization has previously been suggest by Bretthorst (1992).

If the smooth FIR regularization is used as part of an ANCOVA modeling where covariates with no connection to the fMRI response are present, then the regularization term should not be included for those variables

$$\mathbf{R} = \begin{bmatrix} \mathbf{R}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \quad (3.98)$$

Furthermore, if there are several independent inputs (different paradigms) they should be regularized independently, i.e., the regularization matrix \mathbf{R} should be block-diagonal rather than full.

$$\mathbf{R} = \begin{bmatrix} \mathbf{R}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_{22} \end{bmatrix} \quad (3.99)$$

The regularization matrices might be the same $\mathbf{R}_{11} = \mathbf{R}_{22}$.

The posterior density of the parameters $p(\mathbf{b}|\mathbf{y}, \mathbf{X})$ will have a multivariate Gaussian distribution if the noise $\sigma_{\mathbf{u}}^2$ is known (the posterior is conditioned on $\sigma_{\mathbf{u}}^2$) and there is a uniform prior on the parameters: $p(\mathbf{b}) \propto 1$ (Box and Tiao 1992, section 2.7.1). It will still have a Gaussian distribution if a Gaussian prior for $p(\mathbf{b})$ is used (Goutte, Nielsen, and Hansen 2000, equation 6) — the Gaussian prior is a conjugate prior in this connection. However, when $\sigma_{\mathbf{u}}^2$ is not known the distribution of the parameters \mathbf{b} is more complicated: With a uniform prior on \mathbf{b} and $\log \sigma_{\mathbf{u}} \propto 1$ and $p(\sigma_{\mathbf{u}}^2) \propto \sigma_{\mathbf{u}}^{-2}$, respectively — the posterior distribution of \mathbf{b} will follow the multivariate t -distribution (Box and Tiao 1992, section 2.7.2, equation 8.3.13) (Mardia et al. 1979, exercise 2.6.5).

3.16.2 Time-series modeling of fMRI signals

Time-series modeling of the fMRI signal is concerned with the modeling of the interesting signal modulated (filtered) by the hemodynamic response function (HRF), and with modeling of confounding signals. There are several reasons why modeling of the HRF is interesting:

- Understand the underlying physiological models. One of the few models that has been proposed is the balloon (Buxton, Wong, and Frank 1998).
- Obtain a model that can predict the fMRI for new scanning and psychological parameters (so we can optimize them for best detection), e.g., Josephs and Henson (1999) use a HRF to acquire the *estimated measurable power* (EMP) as a function of the stimulus onset asynchrony (SOA) and experiment type by simulation. If the simulation should be valid it is important that the HRF model are correct.
- Obtain a value for the importance of a region in a mental process.

- Extract some features of the response, e.g., the delay. For this the hemodynamic response model should be parametrized not only by its amplitude, but also by, e.g., its delay.
- Identify a hemodynamic response model so the fMRI signal can be deconvolved and the (possibly unknown) neuronal response can be found. This is difficult since the hemodynamic response is not known well enough. In (Højen-Sørensen, Hansen, and Rostrup 1999; Højen-Sørensen, Hansen, and Rasmussen 2000) a Bayesian estimated FIR model for the hemodynamic response was used together with a Markov model for the stimulus.

There has been a few comparison of the different hemodynamic response models: Ollinger, Shulman, and Corbetta (2001a, e.g., figure 4) compared the delayed gamma and the gamma mixture from SPM99 together with the FIR filter (“the F statistic approach”) and find that the gamma provide higher z -values compare to the FIR filter, provided that the paradigm is not shifted. If the paradigm was shifted 1 second then the z -gamma dropped below the FIR filter.

Jittering

In functional neuroimaging experiments — especially event-related fMRI — the timing of the behavioral data (the input) is often different than the brain imaging data, e.g., the sampling rate might be different with the behavioral data being sampled faster and/or the cognitive events might not fall in the same phase as the brain imaging data. It might occur because of variation in reaction time of the subject or because a *jittered* stimulus is used. When jitter is used it becomes possible to interrogate the hemodynamic response at different phases with a higher frequency than the imaging sampling frequency (TR). This is especially useful when the TR are in the order of the characteristic time length of the hemodynamic response.

A way of handling the sample problem would be to upsample the imaging data to the frequency of the behavioral data. This would unfortunately require large amount of storage capacity and it would gain nothing in terms of statistical power. The opposite would be to downsample the behavioral data to TR.

Resampling is a standard exercise in digital signal processing (Mitra 2001, chapter 10), and a few variations are depicted in figure 3.8. The dots denote what would be entered in the design matrix \mathbf{X} if the general linear model was used. The lines shows the model output $\hat{\mathbf{y}}$ if the model was estimated as the linear “Glover” model implemented in SPM (Friston 1999a; Glover 1999). In the lower left corner of the figure is shown a decimation without any filtering at a TR comparable to the event block length. In this case the events fall so no distinction is seen between the varying event block lengths. This simplified downsampling could give rise to false conclusions about the nonlinearity of the response. The right side depicts some more appropriate downsample methods.

A somewhat different approach is possible if the hemodynamic response model is continuous: The SPM toolbox downsamples the behavioral data to a sixteenth of TR and convolves it with the continuous hemodynamic response function with this sample frequency before downsampling fully to TR.

3.16.3 Cross-correlation

Cross-correlation was one of the first methods to be used to analyse fMRI (Bandettini, Jesmanowicz, Wong, and Hyde 1993). In its simplest form the fMRI signal in a pixel is correlated with the paradigm function — usually a boxcar function or a sinusoid. In such a case the cross-correlation can be formulated within the framework of the multiple regression model of equation 3.90 resulting in a t -statistics. A more elaborate cross-correlation will slide the paradigm function across the fMRI signal yielding a *cross-correlation sequence*, see, e.g., (Mitra 2001, section 2.7). The maximum of this function can be reported together with a delay parameter computed as the distance between zero lag and the lag at maximum. A statistics for this situation is difficult to develop analytically, but it has been assessed by Monte Carlo (Goutte and Rostrup 1999).

There are two issues in the computation of the cross-correlation that needs to be considered: When the paradigm is slid across the fMRI signal, the fMRI signal needs to be extended. The usual solution is simple zero-padding. The other issue is how the cross-correlation should be normalized: If the paradigm is in \mathbf{x} and the fMRI signal of a particular voxel is in \mathbf{y} then a “unnormalized cross-correlation” would just be the inner product of the two vectors $r = \mathbf{x}^T \mathbf{y}$ while a “normalized cross-correlation” would standardized the variance in both variables:

$$\rho = \frac{\mathbf{x}^T \mathbf{y}}{\sqrt{\mathbf{x}^T \mathbf{x}} \sqrt{\mathbf{y}^T \mathbf{y}}}. \quad (3.100)$$

In the former case r will be proportional to a b in the multiple regression model and ρ will correspond to a t -statistics.

3.16.4 Delay

When the data represents a time-series and a model is fitted it is sometimes of interest to determine a delay or latency of the response, e.g., the delay of the hemodynamic response in fMRI modeling. There exists different definitions of this

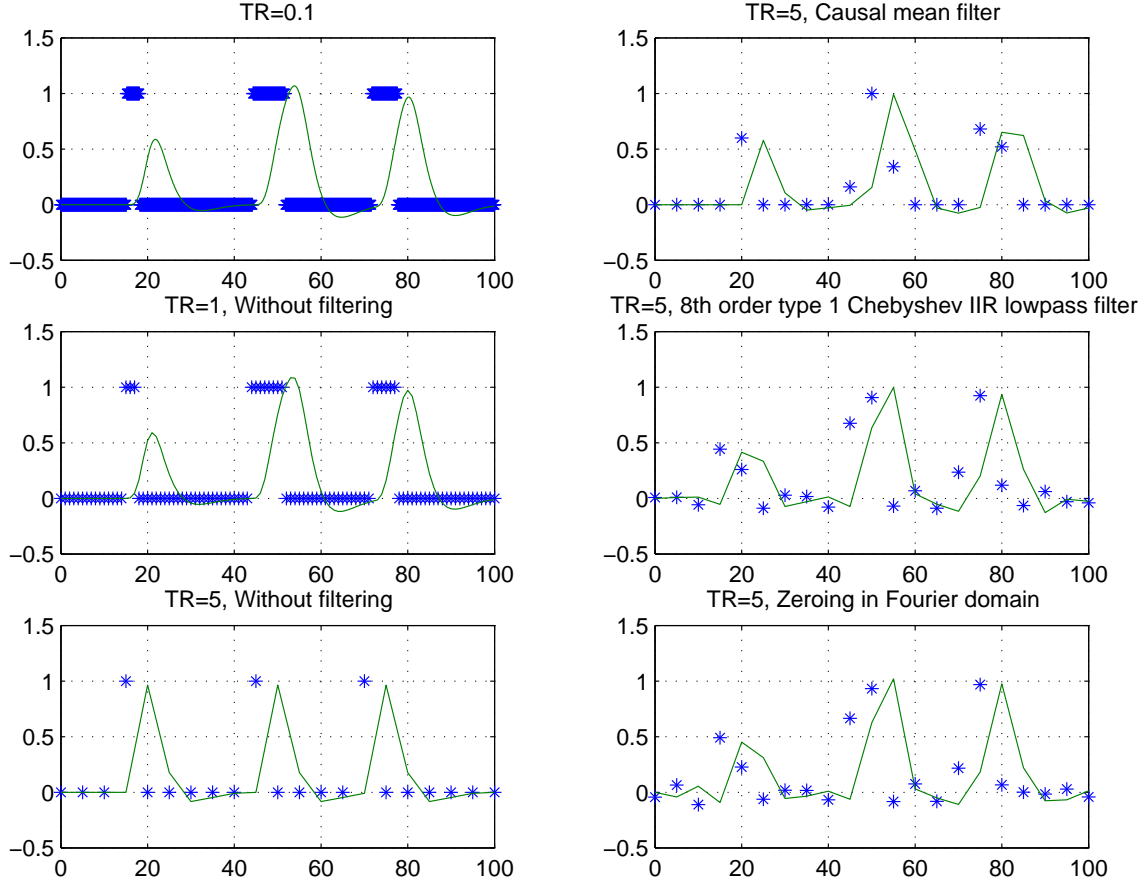


Figure 3.8: Resampling jittered events. Artificial data set with event blocks at 15.2–18.1, 44.3–51.7 and 71.7–77.6 seconds, downsampled to different TRs and convolved with the a gamma density mixture (Friston 1999a). The upper left plot can be considered the truth. The middle right plot is obtained via the standard `decimate.m` function from the Matlab signal processing toolbox, see also (Mitra 2001, section 10.2.4).

- Phase delay is defined as the lag of a sinusoidal signal between the input and output. Thierry, Boulanouar, Kherif, Ranjeva, and Demonet (1999) used this approach according to Liao et al. (2001b)
- The delay can be computed as the *group delay* at zero frequency which corresponds to the *center of mass* of the filter. If b_i represents the FIR filter coefficients the group delay is computed as:

$$\tau = \frac{\sum_i b_i \times i}{\sum_i b_i} \quad (3.101)$$

This delay measure is used in (Goutte et al. 2000; Nielsen et al. 1997).

- For the cross-correlation the delay can be determined as the lag where the cross-correlation function has its maximum. A related delay measure was used in (Bellgowan, Saad, and Bandettini 2001) where the time of maximum of the cross-correlation function between the fitted impulse response function and a gamma density function was used as the delay value. A similar method was used in (Saad, Ropella, Carman, and DeYoe 1996; Saad, Deyoe, and Ropella 1999).
- The peak of the impulse response function can be used as the delay value.
- If the impulse response function is modeled with a parametric model, such as the gamma mixture model then a model parameter might represent a parameter that is appropriate to regard as a delay parameter. This approach was taken by Henson and Rugg (2001) where a gamma density mixture is fitted to an fMRI time-series. It is also implemented in SPM where the first order derivative of the gamma density mixture model (Friston 1999a) can be regarded as a delay parameter.
- The raise time, e.g., the time from 10% to 90% response with a step function as input. Bandettini et al. (1993) used a delay that resembled this measure.

Parameters / Input	Linear	Nonlinear
Linear	GLM with only linear terms	GLM with nonlinear terms
Nonlinear	Gamma convolution model	Neural network

Table 3.16: Examples of nonlinear models: Nonlinearity in parameter-output and in input-output mapping. Inspired by Larsen (1996, table 1.1)

- Menon, Luknowsky, and Gati (1998) regarded an onset time estimate of the delay, where a straight line was fitted to the region between 20 to 70% of peak response and the intercept with zero was taken as the value of the delay.

The two first in the list are often used in the context of digital signal processing, see, e.g., (Mitra 2001, section 4.2.6).

A more elaborate delay estimate is presented by Liao, Worsley, Poline, Duncan, and Evans (2001a, 2001b). This description incorporates inference about the delay parameter.

3.16.5 Random effects

Usually the parameters of the model are regarded as fixed (but unknown). If one of the parameters vary in some systematic way this should be modeled. The former model is known as a *fixed effects model* while the latter is commonly called *random effects model*. If the model consists of both fixed and varying parameter the model is called a *mixed effects model* (Box and Tiao 1992, section 5.1), (Armitage and Berry 1994, section 7.3), (Conradsen 1984a, chapter 5).

The fMRI BOLD response magnitude varies across subjects, see, e.g., (Holmes and Friston 1998). Furthermore, also the intrasubject intersession variation has been observed (McGonigle et al. 2000). With a general linear model (GLM) this can be described with a parameter matrix \mathbf{B} where some of the parameters are stochastic, — vary across subjects and sessions. If the design is *balanced* (e.g., there is an equal amount of scans for each subjects) and parameters are *separable* across the varying component, e.g., separable across subjects, then a two-stage random effects modeling can be employed to estimate the distribution of the stochastic parameter (Holmes and Friston 1998). In this procedure parameter estimates for each subject (or each session) is first identified (in a *first level analysis*), then these are used in a second level analysis in a test, most often a simple *t*-test. Specifically in fMRI analysis a strength parameter can be calculated from, e.g., a gamma density model for each subject and for each voxel. A new data matrix (subject \times voxels) is constructed from the estimated strengths, and this matrix is then analysed with the GLM.

The random effect model is not directly implemented in the Lyngby toolbox, but the two-stage procedure can be applied by making the first stage with a hemodynamic response model, e.g., the FIR model or the “Lange-Zeger”, and the second stage may be run with the ordinary *t*-test. It is not necessary to account for a temporal correlation in the fMRI time-series, but the spatial correlation should still be accounted for if a significance value is required.

3.17 Nonlinear modeling

There are two aspects of nonlinearity in modeling, depending on what effect there is on the output of the model (see also table 3.16):

- **Parameter nonlinearity**, e.g., $\mathbf{Y} = \mathbf{X}\mathbf{f}(\theta)$. When the nonlinearity is in the *parameters* then a linear change in a parameter of the model will result in a nonlinear change in the output with the same input signal. When the input signal is linear changed the output is not necessary nonlinear. Nonlinear models such as these can appear if the input-output relationship is expressed with an ordinary differential equation, where the parameter can be related to the output through an exponential function, while the output is still linear with respect to the input. An example appears in modeling of the fMRI signal where the BOLD response can be modeled with a gamma density function as, e.g., in Lange and Zeger (1997): The parameters that control the shape of the function are nonlinear related with the output and the parameters has to be identified with an iterative scheme; the BOLD response is still modeled as being linear. Another example is the convolution model of (Krugel and von Cramon 1999; Krugel, Zysset, and von Cramon 2000) where the hemodynamic response function is modeled as a Gaussian and the mean and variance parameters (termed lag and dispersion) are nonlinear.
- **Model nonlinearity**, $\mathbf{Y} = \mathbf{f}(\mathbf{X}, \theta)$. On the other hand when the nonlinearity is in the *input-output* mapping, then a linear change in the input will likely result in a nonlinear change in the output. It is still possible that the parameters is linear with respect to the output: $\mathbf{Y} = \mathbf{f}(\mathbf{X})\theta$, — and it is an advantage since it makes the estimation of the parameters easier.

There are many nonlinear models, e.g., Gaussian processes (GP) (Williams and Rasmussen 1996; MacKay 1997), generalized additive models (GAM) (Hastie and Tibshirani 1990) and artificial neural networks (ANN), that will be described below.

3.17.1 Artificial neural networks

The term “artificial neural network” — or just “neural network” — has been used to denote a group of models that are more or less inspired by biological neuronal networks, e.g., linear regression models (such as GLM), self-organizing maps or mixture models. Here the term will solely be used to represent feed-forward neural networks (also called perceptrons) with sigmoidal nonlinearities (activation functions). With a two-layer neural network a model with \mathbf{X} regressed on \mathbf{Y} and with a hyperbolic tangent as the activation function a neural network model can be written as

$$\mathbf{Y} = \tanh(\mathbf{XV})\mathbf{W} + \mathbf{U}, \quad (3.102)$$

where $\mathbf{Y}(N \times Q)$ are observed data and $\mathbf{X}(N \times P)$ is a matrix with either “independent” observed data or design variables. Matrices $\mathbf{V}(P \times R)$ and $\mathbf{W}(R \times Q)$ are model parameters. $\mathbf{U}(N \times Q)$ is a matrix with noise — unobserved random disturbances. In the neural network community \mathbf{X} is called the *input*, $\mathbf{Z} = \tanh(\mathbf{XV})\mathbf{W}$ is the *output*, \mathbf{Y} the *target* and R is the number of *hidden units*. More heterogeneous structures exist where the input is directly connected with the output

$$\mathbf{Y} = \mathbf{XB} + \tanh(\mathbf{XV})\mathbf{W} + \mathbf{U}, \quad (3.103)$$

As with the general linear model the random disturbances are often assumed to be independently Gaussian distributed. The likelihood function for the parameters \mathbf{V} and \mathbf{W} can then be established and the parameters can be optimized for maximum likelihood with any of the methods mentioned in section 3.2.1. For example, with a single output variable $Q = 1$ with Gaussian noise the maximum likelihood estimate for equation 3.102 will be a least squares cost function

$$E_{\text{MLE}} \propto (\mathbf{y} - \hat{\mathbf{y}})^T (\mathbf{y} - \hat{\mathbf{y}}). \quad (3.104)$$

where $\hat{\mathbf{y}} = \tanh(\mathbf{XV})\mathbf{w}$ is the output of the neural network.

It is often found that the number of parameters in the ANN model is large compared to the objects and therefore regularization comes to play an important role in ANN modeling. The setting of the regularization parameter and the determination of the number of hidden units R can be critical and model selection is therefore also important. Ridge-like regularization termed *weight decay* has been suggested (Hinton 1986) as well parameter elimination by methods termed *optimal brain damage* (Le Cun, Denker, and Solla 1990) and *optimal brain surgeon* (Hassibi, Stock, and Wolff 1992). For the weight decay method the cost function of equation 3.104 is simply augmented by the square size of the parameters

$$E_{\text{MAP}} \propto (\mathbf{y} - \hat{\mathbf{y}})^T (\mathbf{y} - \hat{\mathbf{y}}) + \lambda_{\mathbf{V}} (\mathbf{V}^V)^T \mathbf{V}^V + \lambda_{\mathbf{W}} \mathbf{w}^T \mathbf{w}, \quad (3.105)$$

where $\lambda_{\mathbf{V}}$ and $\lambda_{\mathbf{W}}$ are the regularization parameters and “ V ” is a vectorizing operator. Since the augmenting terms can be derived from Gaussian priors on the parameters the cost function of equation 3.105 can be considered a maximum a posteriori cost function.

Optimal brain damage and regularization can be combined (Svarer, Hansen, and Larsen 1993) and this is also implemented in the Lyngby toolbox. If the input dimension P is very large then it can be an advantage to project the data matrix \mathbf{X} to a suitable subspace, e.g., by singular value decomposition or principal component analysis (Lautrup, Hansen, Law, Mørch, Svarer, and Strother 1995), ICA (Mørch 1998) or a canonical analysis (see section 3.18).

Analyzing a neural network — the saliency map

Artificial neural networks are often operated as *black boxes* where no interpretation and no assessment of importance in the predication is given for the individual parameters and the individual inputs (variables). However, pruning and adaptive weight decay implicitly establish the importance of a parameter, and a similar method exists for the input variables with what has been termed a “saliency map” (Baluja and Pomerleau 1995; Mørch, Kjems, Hansen, Svarer, Law, Lautrup, Strother, and Rehm 1995): The map for the importance (for predicting the output) of each single variable in the multivariate input of a model (the neural network).

Subset selection in ordinary linear regression (Mardia et al. 1979, section 6.7) aims at finding a subset of independent variables that will explain most of the variation in the dependent variables. Some of the measures in this connection could potential produce a saliency map, though the multiple correlation coefficient suggest by (Mardia et al. 1979, equation 6.7.1) does not directly generate one single value for each individual input. Another scheme is the so-called delta test (Pi and Peterson 1994; Ohlsson, Peterson, Pi, Rögnavaldsson, and Söderberg 1994).

When the parameters of the model (the neural network) have associated individual regularization parameters the adapted parameters might be taken as the saliency map: Inputs associated with parameters with large regularization are given a small saliency. In Bayesian settings this has been called *automatic relevance determination* (MacKay 1994; MacKay 1995a; Neal 1996, sections 1.2.3 and 4.3)

The saliency scheme taken in (Mørch, Kjems, Hansen, Svarer, Law, Lautrup, Strother, and Rehm 1995; Mørch 1998; Mørch, Hansen, Law, Strother, Svarer, Lautrup, Kjems, Lange, and Paulson 1996a) is related to the optimal brain damage technique. The basic idea is to evaluate the increase in the generalization error when a input variable is deleted. As it is infeasible to re-estimate the parameters of the model for each individual input variable when the number of input variables is large, a loss in generalization error is assessed by using a Taylor expansion of the generalization error.

3.17.2 Nonlinear modeling in functional neuroimaging

There are a number of ways in which nonlinear models can be applied in functional neuroimaging:

- Nonlinear regression for parametric activation designs. Consider a saccadic eye movement study, where subjects are moving their eyes toward an external visual stimuli at varying frequency. There might exist a nonlinear relationship between the frequency of the eye movement and the measured brain signal, e.g., the rCBF. The task of the nonlinear regression is to model this relationship.

This kind of modeling can be made within the general linear model by augmenting the design matrix with nonlinear terms, e.g., in connection with a saccadic eye movement the frequency would be in one column of the design matrix and in another column the squared frequency which should model the nonlinear effect. Such an approach is taken in Büchel et al. (1996) and nonlinear effects are demonstrated on PET data sets with second order polynomial expansion and cosine basis functions. Büchel et al. (1998) describes the similar method on an fMRI study with a block design (34 seconds activations — 34 seconds baseline): The design matrix has three columns with the zeroth, first and second order polynomial expansion, and the values in the design matrix corresponding to the baseline are zero, while the values corresponding to activation scans are non-zero.

This kind of nonlinear modeling can also be performed with linear models if “agnostic labeling” is used for each different parameter value. Then there has to be one column in the design matrix for each parameter value, which gives a large DOF with less power in a statistical test, but is more flexible.

- Nonlinear time effects for efMRI. In event-related fMRI (efMRI) it has been shown that rapidly presented events produce a nonlinear BOLD response, see section 2.3.4.

Friston, Josephs, Rees, and Turner (1998b) and Friston, Mechelli, Turner, and Price (2000b) applied Volterra-series expansion based on the Gamma mixture density hemodynamic response function. ANN model as represented in equation 3.102 was described shortly by (Nielsen, Goutte, and Hansen 1999). ANN models were also used by (Pedersen 1997; Lange, Hansen, Pedersen, Savoy, and Strother 1996; Knowles, Manton, Issa, and Turnbull 1998a; Knowles, Manton, and Turnbull 1998b). The *Balloon model* (Buxton, Wong, and Frank 1998) represents a plausible physiological model and the parameters in this model are non-linear.

- Design variables can couple nonlinearly to individual voxel activations. This is interesting when the cognitive components are not additive and can be modeled within the GLM by including an interaction term (Friston, Holmes, Worsley, Poline, Frith, and Frackowiak 1995d), e.g., by multiplying two design variables to form a third variable.
- Voxels activations couple nonlinearly to a design variable. This cannot be handled with the ordinary GLM. The unsupervised nonlinear PCA method (Friston, Philips, Chawla, and Büchel 1999b; Friston, Philips, Chawla, and Büchel 2000c) is able to model nonlinearities with second order interactions. An advantage of this model is that linear and nonlinear effects are modeled in different parameters, enabling an identification of interaction effects. Another method of dealing with this kind of nonlinearity is by using nonlinear regression from the brain scans data to the design variables. Using neural network this has been developed by the group of Lars Kai Hansen, see, e.g., (Mørch et al. 1997; Lundsager and Kristensen 1996; Mørch et al. 1996b). Contrary to the nonlinear PCA method it does not separate linear and interaction effects.

3.18 Canonical analyses

“Canonical analyses” (as it will be termed here) concern the description of the relationship between (a set of) two multivariate variables \mathbf{X} and \mathbf{Y} , e.g., to *summarize their relationship in a subspace* or to *predict one from the other*. The *linear analyses* of this type can be said to include *canonical correlation analysis*, *partial least squares*, *canonical ridge analysis* and *principal component regression* among others. Some of these can be regarded as “bottleneck models” where the relation between the two sets of variables is through a low dimensional subspace, hence the name “low-dimensional linear models” (Borga 1998). In other case the \mathbf{X} and \mathbf{Y} are regarded as being related to (generated from) a hidden subspace or *latent variables*, hence “latent variable models”, see, e.g., (Burnham, MacGregor, and Viveros 1999). The type of models can also be considered as “heteroassociative models” (Diamantaras and Kung 1996, section 6). Other names are bilinear models (BLM, Martens and Næs 1989) and reduced rank models.

Usually the variables are considered Gaussian distributed, but models for multinomial data via a softmax layer have also been described (Yee and Hastie 2000).

3.18.1 Canonical correlation analysis

Canonical correlation analysis (CCA) was developed in (Hotelling 1935; Hotelling 1936). Introductions to CCA are available in Mardia et al. (1979, chapter 10) and (Anderson 1984, chapter 12). The estimation problem in CCA is to find two matrices $\mathbf{A}(P \times K)$ and $\mathbf{B}(Q \times K)$ so that the correlations between the linear projects $\Phi = \mathbf{X}\mathbf{A}$ and $\Psi = \mathbf{Y}\mathbf{B}$ is the

largest. To identify this uniquely the solution should be constrained, and the typical constraints are that the projections $\Phi(N \times K)$ and $\Psi(N \times K)$ should be orthogonal $\Phi^T \Phi = \Psi^T \Psi = \mathbf{I}_K$ and the elements in Φ and Ψ are to be arranged so that the first column pair η_1 and ψ_1 has the highest correlation, the second column pair the second highest correlation, and so forth.

There are different ways of estimating the parameters \mathbf{A} and \mathbf{B} with Mardia et al. (1979) relying on SVD:

$$\mathbf{K} = \mathbf{S}_{xx}^{-1/2} \mathbf{S}_{xy} \mathbf{S}_{yy}^{-1/2}, \quad (3.106)$$

where covariances are set as $\mathbf{S}_{xx} = N^{-1} \mathbf{X}^T \mathbf{H} \mathbf{X}$, $\mathbf{S}_{yy} = N^{-1} \mathbf{Y}^T \mathbf{H} \mathbf{Y}$ and $\mathbf{S}_{xy} = N^{-1} \mathbf{X}^T \mathbf{H} \mathbf{Y}$ with \mathbf{H} being the centering matrix.

$$\mathbf{ULV}^T = \text{svd}[\mathbf{K}] \quad (3.107)$$

$$\mathbf{A} = \mathbf{S}_{xx}^{-1/2} \mathbf{U} \quad (3.108)$$

$$\mathbf{B} = \mathbf{S}_{yy}^{-1/2} \mathbf{V}, \quad (3.109)$$

and with a perhaps abusive notation:

$$\mathbf{ALB}^T = \mathbf{S}_{xx}^{-1/2} \text{svd} \left[\mathbf{S}_{xx}^{-1/2} \mathbf{S}_{xy} \mathbf{S}_{yy}^{-1/2} \right] \mathbf{S}_{yy}^{-1/2} \quad (3.110)$$

The diagonal elements of \mathbf{L} are called the *canonical correlation coefficients*, and Mardia et al. (1979) call the columns of \mathbf{A} and \mathbf{B} the *canonical correlation vectors*. Furthermore, Φ and Ψ are called the *canonical correlation variables*. If one of the matrices, e.g., $\mathbf{Y}(N \times Q) = \mathbf{Y}(N \times (K - 1))$ is an indicator matrix indicating group membership in K groups for the N objects in \mathbf{X}

$$y_{nk} = \begin{cases} 1 & \text{if } \mathbf{x}_n \text{ in the } k\text{'th group,} \\ 0 & \text{otherwise,} \end{cases} \quad (3.111)$$

then (the first column of) \mathbf{H} are also called *Fisher's linear discriminant function* (Fisher 1936) or the *first canonical variate* (Mardia et al. 1979, exercise 11.5.4), — with the terminology from discriminant analysis. The first column of \mathbf{A} (\mathbf{a}_1) is often derived as the maximum eigenvector from a (generalized) eigenvalue decomposition:

$$\mathbf{a}_1 = \max \text{eig} \left[\mathbf{W}^{-1} \tilde{\mathbf{B}} \right], \quad (3.112)$$

where \mathbf{W} and $\tilde{\mathbf{B}}$ are sum of squares and product (SSP) matrices:

$$\mathbf{T} = \mathbf{W} + \tilde{\mathbf{B}} \quad \text{Total SSP matrix,} \quad (3.113)$$

$$\mathbf{W} = N_{k=1} \mathbf{S}_{xx, k=1} + \dots + N_{k=K} \mathbf{S}_{xx, k=K} \quad \text{Within group SSP matrix,} \quad (3.114)$$

$$\tilde{\mathbf{B}} = \mathbf{T} - \mathbf{W} \quad \text{Between group SSP matrix.} \quad (3.115)$$

Such an analysis are often referred to as Fisher's linear discriminant analysis (FLD/LDA), canonical variate analysis (CVA) or canonical discriminant analysis (CDA).

This ordinary canonical correlation analysis will not work if any of the two matrices \mathbf{X} and \mathbf{Y} are singular, since the covariance matrices \mathbf{S}_{xx} and \mathbf{S}_{yy} will not be invertable. The canonical correlation coefficients are not affected by non-singular affine transformations and the canonical correlation vectors are related through a linear transformation (Mardia et al. 1979, theorem 10.2.4). Furthermore the canonical correlation vectors are not (necessarily) orthogonal.

3.18.2 Partial least squares

"Partial least squares" (PLS) constitutes a group of algorithms and model, see table 3.17 for some of them. Canonical correlation analysis can be regarded as a *mode B PLS* algorithm (Martens and Næs 1989, page 165). The one considered here is where the parameters of the model are determined through singular value decomposition of the covariance matrix between \mathbf{X} and \mathbf{Y} (McIntosh et al. 1996; Worsley et al. 1997) — a decomposition associated with the SIMPLS algorithm (de Jong 1993)

$$\mathbf{ULV}^T = \text{svd}(\mathbf{X}^T \mathbf{Y}) \quad (3.116)$$

The "canonical vectors" \mathbf{U} and \mathbf{V} can of course also be found through eigenvalue decomposition of the inner and outer product matrix (equations 3.22 and 3.23). It should be immediately apparent that if one of the matrices is the identity matrix then PLS will correspond to ordinary SVD on the other matrix. Unlike canonical correlation analysis the PLS estimate depends on the centering and scaling of the data in \mathbf{X} and \mathbf{Y} matrices and if no prior knowledge is available about the data it is recommend to scale and center the variables before the singular value decomposition (Wold, Ruhe,

Abbr.	Description	Reference
NIPALS	“Nonlinear iterative PLS”	Wold (1966)
N-PLS	“N-Way PLS”, \mathbf{X} is of size $N \times P \times Q \times \dots$	
PLS1	One-dimensional “response” matrix: $\mathbf{Y}(N \times 1) = \mathbf{y}$	
PLS2	Multidimensional “response” matrix: $\mathbf{Y}(N \times Q)$, $Q > 1$	e.g., Martens and Næs (1989, page 146+)
PLS A mode	Regression from latent variables to measured	e.g., Martens and Næs (1989, page 188)
PLS B mode	Regression from measured variables to latent	e.g., Martens and Næs (1989, page 188)
PLS C mode	Combination of A and B mode	e.g., Martens and Næs (1989, page 188)
SIMPLS	Eigendecomposition of covariance matrix	de Jong (1993)

Table 3.17: Some PLS algorithms and models. See also (de Jong and Phatak 1997) for a description of some of the PLS types.

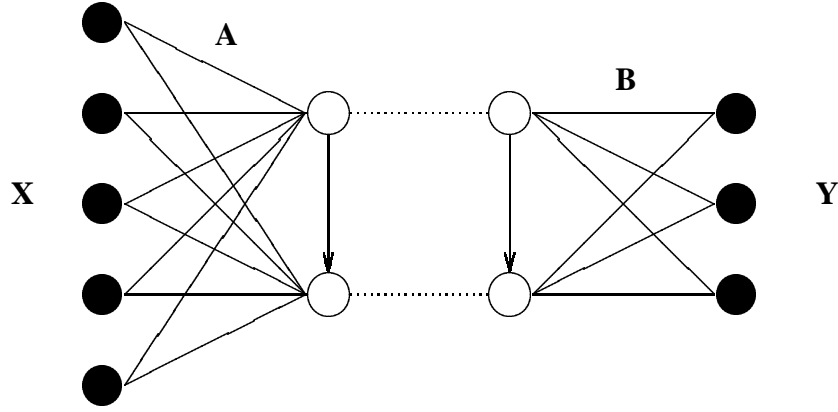


Figure 3.9: Cross-correlation APCA network between $\mathbf{X}(N \times P)$ and $\mathbf{Y}(N \times Q)$ with $P = 5$ and $Q = 3$ with lateral orthogonization between the $K = 2$ hidden units. Black circles denote directly observed data. Comparable to (Diamantaras and Kung 1996, figure 6.5(a))

Wold, and Dunn III 1984, section 5), e.g., here with a $N \times N$ centering matrix \mathbf{H} and diagonal matrices \mathbf{D}_x and \mathbf{D}_y containing the square root of the diagonal elements in the covariance matrices of \mathbf{X} and \mathbf{Y}

$$\mathbf{ULV}^T = \text{svd}(\mathbf{D}_x^{-1} \mathbf{X}^T \mathbf{H}_N \mathbf{Y} \mathbf{D}_y^{-1}). \quad (3.117)$$

The PLS method is somewhat similar to the *cross-correlation asymmetric principal component analysis* (cross-correlation APCA) (Diamantaras and Kung 1994a) and (Diamantaras and Kung 1996, section 6.5)⁵. Diamantaras and Kung (1996, section 6.5) developed the estimation in the model shown by the connectionistic network in figure 3.9 and the maximization of the following cost function

$$E_{\text{APCA}} = \text{trace}(\mathbf{A} \mathbf{S}_{xy} \mathbf{B}^T), \quad (3.118)$$

with the orthogonality constraint on the parameter matrices $\mathbf{A}(K \times P)$ and $\mathbf{B}(K \times Q)$

$$\mathbf{A}^T \mathbf{A} = \mathbf{B}^T \mathbf{B} = \mathbf{I}. \quad (3.119)$$

As cross-correlation APCA does not put constraints on the “ordering” of the vectors in \mathbf{A} and \mathbf{B} any solution can be rotated with an orthogonal matrix.

The parameters in the cross-correlation APCA model can be found through an iterative scheme called “cross-coupled Hebbian rule”.

⁵Diamantaras and Kung (1996, equation 6.27) approximately defines “Cross-correlation” as the inner product of the two matrices: $N^{-1} \mathbf{X}^T \mathbf{Y}$ (not considering centering of the matrices). “Correlation” is often used differently in statistics, digital signal processing and the neural network literature depending on whether it is normalized with the standard deviation of the individual variables or not and whether it is centered. $N^{-1} \mathbf{D}^{-1} \mathbf{X}^T \mathbf{H} \mathbf{X} \mathbf{D}^{-1}$ is denoted the (sample) “correlation matrix” by Mardia, Kent, and Bibby (1979, equation 1.4.13) and Conradsen (1984b, page 2.6) while $N^{-1} \mathbf{X}^T \mathbf{X}$ is denoted the “correlation matrix” by (Hertz, Krogh, and Palmer 1991, equation 8.4) and Diamantaras and Kung (1996, equation 6.27). In digital signal processing cross-correlation is often regarded as between two signal in time where the cross-correlation in lag zero would be defined as $\mathbf{x}^T \mathbf{y}$ or $N^{-1} \mathbf{x}^T \mathbf{y}$ depending on the signal, see, e.g., (Mitra 2001, equations 2.103 and 2.144). See also note by Borga (1998, page 25).

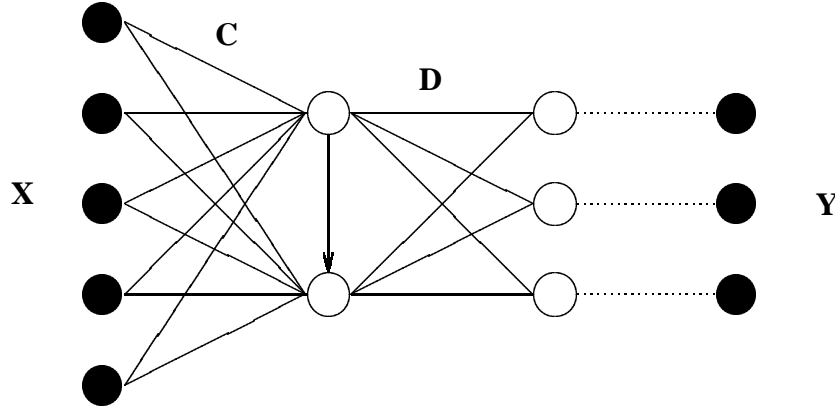


Figure 3.10: Linear approximation APCA network between input $\mathbf{X}(N \times P)$ and target $\mathbf{Y}(N \times Q)$ with $P = 5$ and $Q = 3$ lateral orthogonalization between the $K = 2$ hidden units. The black circles are variables directly observed while the white circles are hidden variables. \mathbf{B} from equation 3.121 is $\mathbf{B} = \mathbf{C}\mathbf{D}$. The network structure is different from (Diamantaras and Kung 1996, figure 6.1): An extra layer has been added to incorporate the “noise” between $\mathbf{X}\mathbf{B}$ and \mathbf{Y} .

3.18.3 Orthonormalized PLS, redundancy analysis, etc.

“Orthonormalized partial least squares” (OPLS) is proposed by Worsley et al. (1997) where one of the matrices is normalized by the square root of the full covariance matrix.

$$\tilde{\mathbf{U}}\mathbf{L}\mathbf{V}^T = \text{svd} \left(\mathbf{S}_{\mathbf{xx}}^{-1/2} \mathbf{X}^T \mathbf{Y} \right). \quad (3.120)$$

In (Worsley 1997; Worsley et al. 1997) CCA, PLS and OPLS are called *canonical variables analyses*. The *linear heteroencoder with bottleneck* (LHB) as described by Roweis and Brody (1999, section 3) uses the same subspace as OPLS for its singular value decomposition. The bottleneck is shown to be optimal with the following cost function — the *reconstruction error*

$$E_{\text{LHB}} = \|\mathbf{X}\mathbf{B} - \mathbf{Y}\|_F^2 = \text{trace} \left[(\mathbf{X}\mathbf{B} - \mathbf{Y})^T (\mathbf{X}\mathbf{B} - \mathbf{Y}) \right], \quad (3.121)$$

and restricting the rank of the parameters $K = \text{rank}(\mathbf{B})$ to a number less than $\min[\text{rank}(\mathbf{X}), \text{rank}(\mathbf{Y})]$

$$\hat{\mathbf{B}}_{\text{LHB},K} = (\mathbf{X}^T \mathbf{X})^{-1/2} \text{svd}_K \left[(\mathbf{X}^T \mathbf{X})^{-1/2} \mathbf{X}^T \mathbf{Y} \right] \quad (3.122)$$

where svd_K is the truncated singular value decomposition to rank K .

Similar models have been studied by several authors: According to Jackson (1991, section 12.6.4) Rao (1964) proposed it as *principal components of instrumental variables* and Fortier (1966) introduced the same model. Van den Wollenberg (1977) called it redundancy analysis (RDA). Diamantaras and Kung (1996, chapter 6) account that Scharf (1991) introduced it as *reduced-rank Wiener filtering* and that Baldi and Hornik (1989) also studied it. Diamantaras and Kung (1996) and Diamantaras and Kung (1994b) themselves called it *linear approximation asymmetric principal component analysis*, the network structure is depicted in figure 3.10. Burnham, MacGregor, and Viveros (1999) referred to it as *reduced-rank regression* (RRR). Furthermore, Borge (1998) has denoted it *multivariate linear regression* (MLR).

Diamantaras and Kung (1996) show that parameter estimation can also be obtained by generalized singular value decomposition (GSVD) of a matrix pair $\mathbf{X}^T \mathbf{Y}$ and \mathbf{X} , and (Borge 1998, section 4.5) describes a solution based on generalized eigenvalue decomposition.

RRR is one of the end points — with PCR on the other end — in the so-called *joint continuum regression* that is able to interpolate between the two methods on a continuous scale by adjusting a single parameter (Brooks and Stone 1994).

As a regression model RRR is optimal and can perform better than PCR, PLS and CCA (Jackson 1991, table 12.8, Linnerud data) if the performance is measured as the trace of the covariance to the residuals. If the performance is measured by the determinant of the residual covariance the CCA is the best.

3.18.4 Canonical ridge analysis

Canonical ridge analysis (CRA) (Vinod 1976) can be viewed as canonical correlation analysis with regularization, and as with ordinary ridge regression small non-zero regularization parameters helps to stabilize the solution (Mardia et al. 1979, exercise 10.2.14). Canonical ridge analysis adds a diagonal term to the $\mathbf{S}_{\mathbf{xx}}$ and $\mathbf{S}_{\mathbf{yy}}$ covariance matrices

$$\tilde{\mathbf{S}}_{\mathbf{xx}} = \mathbf{S}_{\mathbf{xx}} + k_X \mathbf{I} \quad (3.123)$$

$$\tilde{\mathbf{S}}_{\mathbf{yy}} = \mathbf{S}_{\mathbf{yy}} + k_Y \mathbf{I}, \quad (3.124)$$

and then CRA proceeds as ordinary CCA with these regularized covariance matrices. The regularization parameters k_X and k_Y should be non-negative, and if they are positive $k_X, k_Y > 0$ then the regularized covariance $\tilde{\mathbf{S}}$ will be invertable even if \mathbf{S} are singular.

With a slightly different application of the regularization parameters the regularized covariance can be written as

$$\tilde{\mathbf{S}}_{xx} = (1 - k_X)\mathbf{S}_{xx} + k_X\mathbf{I} \quad (3.125)$$

$$\tilde{\mathbf{S}}_{yy} = (1 - k_Y)\mathbf{S}_{yy} + k_Y\mathbf{I}, \quad (3.126)$$

with $0 \leq k_X, k_Y \leq 1$. Ordinary CCA are obtained with $k_X = k_Y = 0$ and the PLS of the previous section can be obtained with $k_X = k_Y = 1$. Thus CRA forms a continuum between PLS and CCA as extremes (Nielsen, Hansen, and Strother 1998). It can be noted that if a matrix is orthonormal implying $\mathbf{S} = \mathbf{I}$ then the regularization for that matrix has no effect. This could be the case in connection with a design matrix.

This type of regularization has also been used in discriminant analysis and termed penalized discriminant analysis (PDA), where it is usually formulated as a regularization on the within-group SSP matrix (Ripley 1996, equation 3.8)

$$\tilde{\mathbf{W}} = \mathbf{W} + k_W\mathbf{I} \quad (3.127)$$

or the total SSP matrix (Kustra and Strother 2000). Krzanowski, Jonathan, McCarthy, and Thomas (1995, section 3.1) referred to this as *generalized ridge discrimination* (GRD).

3.18.5 Canonical correlation analysis with singular matrices

Ordinary canonical correlation analysis is not possible if any of the two matrices \mathbf{X} and \mathbf{Y} are singular because it will not be possible to invert the covariances \mathbf{S}_{xx} and \mathbf{S}_{yy} . It has been suggested to use generalized inverses (Mardia, Kent, and Bibby 1979, page 287), e.g., the Moore-Penrose inverse. Krzanowski, Jonathan, McCarthy, and Thomas (1995) term a related method in discriminant analysis based on the generalized inverse for *modified canonical analysis* (MCA). However, this will not identify the canonical vectors uniquely for canonical correlation analysis with $Q > 1$ as all structure is taken out of the matrices so that $\mathbf{K} = \mathbf{S}_{xx}^{-1/2}\mathbf{S}_{xy}\mathbf{S}_{yy}^{-1/2}$ will be spheric (all non-zero singular values are unity), and it will not be possible to identify a “first” canonical component due to the rotational ambiguity. A subspace can be identified however.

The use of SVD or PCA (so-called preliminary principal component analysis) will correspond to using the Moore-Penrose inverse if all components are used, but if only some of them are used it will be possible to identify the singular vectors uniquely. First the \mathbf{X} and \mathbf{Y} matrix are decomposed via SVD

$$\mathbf{ULV}^T = \text{svd}[\mathbf{X}] \quad (3.128)$$

$$\mathbf{CDE}^T = \text{svd}[\mathbf{Y}]. \quad (3.129)$$

Then the principal components $\tilde{\mathbf{X}}$ and $\tilde{\mathbf{Y}}$ are obtained with L and M as the number of components maintained in the analysis

$$\tilde{\mathbf{X}}_L = \mathbf{U}_L\mathbf{L}_L \quad (3.130)$$

$$\tilde{\mathbf{Y}}_M = \mathbf{C}_M\mathbf{D}_M. \quad (3.131)$$

The principal component are used as “input” to CCA and intermediate vectors are obtained

$$\mathbf{PRQ}^T = \text{svd}[\tilde{\mathbf{K}}] = \text{svd}[\mathbf{S}_{\tilde{x}\tilde{x}}^{-1/2}\mathbf{S}_{\tilde{x}\tilde{y}}\mathbf{S}_{\tilde{y}\tilde{y}}^{-1/2}] \quad (3.132)$$

$$= \text{svd}[\mathbf{U}_L^T\mathbf{C}_M]. \quad (3.133)$$

Finally the canonical correlation vectors $\mathbf{A}(P \times K)$ and $\mathbf{B}(Q \times K)$ in the space of \mathbf{X} and \mathbf{Y} is found

$$\mathbf{A} = \mathbf{V}_L\mathbf{L}_L^{-1}\mathbf{P} \quad (3.134)$$

$$\mathbf{B} = \mathbf{E}_M\mathbf{D}_M^{-1}\mathbf{Q} \quad (3.135)$$

Another possibility is to use CRA to make the covariance matrices \mathbf{S}_{xx} and \mathbf{S}_{yy} positive definite. The ridge regularization can be applied in the in the full space or in the space spanned by the principal components from a preliminary principal component analysis (PPCA), — the result does not differ, provided that all PPCA components are maintained and that the principal component maintain the variance, i.e., not sphered The canonical correlation vectors can then be found through a low-dimensional eigenvalue decomposition, see theorem B.2.1 on page 146

$$\mathbf{A} = \mathbf{V}\check{\mathbf{\Lambda}}_x\mathbf{U}^T\check{\mathbf{P}} \quad (3.136)$$

$$\mathbf{B} = \mathbf{E}\check{\mathbf{\Lambda}}_y\mathbf{C}^T\check{\mathbf{Q}} \quad (3.137)$$

$$\mathbf{R} = \check{\mathbf{\Psi}}^{1/2} = \check{\mathbf{\Upsilon}}^{1/2}, \quad (3.138)$$

where $\check{\mathbf{P}}(N \times K)$ and $\check{\mathbf{Q}}(N \times K)$ are scaled right eigenvectors and the diagonal matrices $\Psi(K \times K)$ and $\Upsilon(K \times K)$ contain K eigenvalues each from a symmetric $K \times K$ eigenvalue decomposition

$$\mathbf{P}\Psi\mathbf{P}^{-1} = \text{reig}(\mathbf{C}\Lambda_y^2\mathbf{C}^T\mathbf{U}\Lambda_x^2\mathbf{U}^T) \quad (3.139)$$

$$\mathbf{Q}\Upsilon\mathbf{Q}^{-1} = \text{reig}(\mathbf{U}\Lambda_x^2\mathbf{U}^T\mathbf{C}\Lambda_y^2\mathbf{C}^T) \quad (3.140)$$

$$\check{\mathbf{P}} = \mathbf{P} \left\{ \text{diag} [\text{diag} (\mathbf{P}^T\mathbf{U}\Lambda_x^2\mathbf{U}^T\mathbf{P})] \right\}^{-1/2} \quad (3.141)$$

$$\check{\mathbf{Q}} = \mathbf{Q} \left\{ \text{diag} [\text{diag} (\mathbf{Q}^T\mathbf{C}\Lambda_y^2\mathbf{C}^T\mathbf{Q})] \right\}^{-1/2} \quad (3.142)$$

$$\check{\Lambda}_x = (\mathbf{L}^2 + k_x\mathbf{I}_K)^{-1}\mathbf{L} \quad (3.143)$$

$$\check{\Lambda}_y = (\mathbf{D}^2 + k_y\mathbf{I}_K)^{-1}\mathbf{D} \quad (3.144)$$

$$\Lambda_x = (\mathbf{L}^2 + k_x\mathbf{I}_K)^{-1/2}\mathbf{L} \quad (3.145)$$

$$\Lambda_y = (\mathbf{D}^2 + k_y\mathbf{I}_K)^{-1/2}\mathbf{D}, \quad (3.146)$$

where \mathbf{B} may also be computed as

$$\mathbf{B} = \mathbf{E}\check{\Lambda}_y\mathbf{C}^T\mathbf{U}\Lambda_x^2\mathbf{U}^T\check{\mathbf{P}} \quad (3.147)$$

3.18.6 Bilinear modeling

Martens and Næs (1989, section 3.3.3) and Burnham, MacGregor, and Viveros (1999) regard the so-called bilinear model (BLM) particularly used in chemometrics

$$\mathbf{X} = \mathbf{T}\mathbf{P} + \mathbf{E} \quad (3.148)$$

$$\mathbf{Y} = \mathbf{T}\mathbf{Q} + \mathbf{F}, \quad (3.149)$$

where $\mathbf{X}(N \times P)$ and $\mathbf{Y}(N \times Q)$ are measured variables, $\mathbf{T}(N \times K)$ contains latent variables, $\mathbf{P}(K \times P)$ and $\mathbf{Q}(K \times Q)$ are parameter matrices (“regression coefficients”) and $\mathbf{E}(N \times P)$ and $\mathbf{F}(N \times Q)$ are residuals (random errors, noise, ...). This model can be regarded as a PLS A mode model (see table 3.17) with the observed variables as a function of the latent variables. It is often presented in a regression framework where the task is to predict \mathbf{Y} from \mathbf{X} where it aims to substitute the multivariate regression model $\mathbf{Y} = \mathbf{X}\mathbf{B} + \mathbf{U}$ and its maximum likelihood (ordinary least squares) estimate $\hat{\mathbf{B}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}$ with an estimate that is better for rank-deficient and ill-posed problems. This is done by making K smaller than the rank of $\mathbf{X}^T\mathbf{X}$.

One way to estimate the parameters in this model is to estimate the latent variables \mathbf{T} from the observed variables \mathbf{X} , (Martens and Næs 1989, equation 3.34), (Burnham, MacGregor, and Viveros 1999, equation 3)

$$\hat{\mathbf{T}} = \mathbf{X}\Phi. \quad (3.150)$$

This model is not uniquely determined, and a common constraint for the latent variables is an orthonormality constraint

$$\hat{\mathbf{T}}^T\hat{\mathbf{T}} = \mathbf{I}_K. \quad (3.151)$$

If the noise matrices are independent Gaussian distributed $\mathbf{e} \sim \mathcal{N}_P(\mathbf{0}, \Sigma_{ee})$ and $\mathbf{f} \sim \mathcal{N}_Q(\mathbf{0}, \Sigma_{ff})$ then the joint log-likelihood can be written as (Burnham, MacGregor, and Viveros 1999, equation 6)

$$\ell = -\frac{1}{2}\text{trace}[(\mathbf{X} - \mathbf{X}\Phi\mathbf{P})\Sigma_{ee}^{-1}(\mathbf{X} - \mathbf{X}\Phi\mathbf{P})^T] - \frac{1}{2}\text{trace}[(\mathbf{Y} - \mathbf{X}\Phi\mathbf{Q})\Sigma_{ff}^{-1}(\mathbf{Y} - \mathbf{X}\Phi\mathbf{Q})^T] \quad (3.152)$$

The profile log-likelihood for Φ can be found when maximum likelihood estimates for \mathbf{P} and \mathbf{Q} with fixed Φ are plugged into equation 3.152 (Burnham, MacGregor, and Viveros 1999, equation 9)

$$\ell(\Phi) = \frac{1}{2}\text{trace}[\Phi^T(\mathbf{X}^T\mathbf{X}\Sigma_{ee}^{-1}\mathbf{X}^T\mathbf{X} + \mathbf{X}^T\mathbf{Y}\Sigma_{ff}^{-1}\mathbf{Y}^T\mathbf{X})\Phi] \quad (3.153)$$

The maximum likelihood estimate of the parameter Φ is found as the eigenvectors to a generalized eigenvalue problem

$$(\mathbf{X}^T\mathbf{X}\Sigma_{ee}^{-1}\mathbf{X}^T\mathbf{X} + \mathbf{X}^T\mathbf{Y}\Sigma_{ff}^{-1}\mathbf{Y}^T\mathbf{X})\Phi = \mathbf{X}^T\mathbf{X}\Phi\mathbf{D} \quad (3.154)$$

Burnham, MacGregor, and Viveros (1999) show that by varying the covariance of the residuals the estimate of Φ would vary between an estimate associated with principal component regression (PCR) and another associated with redundancy analysis (RDA). This is done by assuming uncorrelated noise $\Sigma_{ee} = \sigma_e^2\mathbf{I}$ and $\Sigma_{ff} = \sigma_f^2\mathbf{I}$

$$(\sigma_e^2\mathbf{X}^T\mathbf{X}\mathbf{X}^T\mathbf{X} + \sigma_f^2\mathbf{X}^T\mathbf{Y}\mathbf{Y}^T\mathbf{X})\Phi = \mathbf{X}^T\mathbf{X}\Phi\mathbf{D} \quad (3.155)$$

By multiplication with $\sigma_e \sigma_f$, setting $\phi = \sigma_e / \sigma_f$ and $\check{\mathbf{D}} = \sigma_e \sigma_f \mathbf{D}$

$$\left(\frac{1}{\phi} \mathbf{X}^\top \mathbf{X} \mathbf{X}^\top \mathbf{X} + \phi \mathbf{X}^\top \mathbf{Y} \mathbf{Y}^\top \mathbf{X} \right) \Phi = \mathbf{X}^\top \mathbf{X} \Phi \check{\mathbf{D}}. \quad (3.156)$$

With $\phi \rightarrow \infty$ corresponding to the errors in \mathbf{X} being much larger than those of \mathbf{Y} the estimate associated with redundancy analysis (RDA, section 3.18.3) is obtained

$$(\mathbf{X}^\top \mathbf{X} \mathbf{X}^\top \mathbf{Y} \mathbf{Y}^\top \mathbf{X}) \Phi = \Phi \mathbf{D}. \quad (3.157)$$

$$\mathbf{S}_{\mathbf{xx}}^{-1} \mathbf{S}_{\mathbf{xy}} \mathbf{S}_{\mathbf{yx}} \Phi = \Phi \mathbf{D}, \quad (3.158)$$

Here the mean is disregarded in the calculation of the covariances. With $\phi \rightarrow 0$ corresponding to errors in \mathbf{Y} being much larger than those of \mathbf{X} the estimate is associated with principal component regression (PCR)

$$(\mathbf{X}^\top \mathbf{X}) \Phi = \Phi \mathbf{D}. \quad (3.159)$$

$$\mathbf{S}_{\mathbf{xx}} \Phi = \Phi \mathbf{D}. \quad (3.160)$$

Similarly, the left canonical correlation vector can be obtained by setting the covariances to $\Sigma_{\mathbf{ee}} = \mathbf{X}^\top \mathbf{X}$ and $\Sigma_{\mathbf{ff}} = \mathbf{Y}^\top \mathbf{Y}$, ignoring the normalization term N^{-1}

$$(\mathbf{S}_{\mathbf{xx}} \mathbf{S}_{\mathbf{xx}}^{-1} \mathbf{S}_{\mathbf{xx}} + \mathbf{S}_{\mathbf{xy}} \mathbf{S}_{\mathbf{yy}}^{-1} \mathbf{S}_{\mathbf{yx}}) \Phi = \mathbf{S}_{\mathbf{xx}} \Phi \mathbf{D} \quad (3.161)$$

With an invertable $\mathbf{S}_{\mathbf{xx}}$ and with $\tilde{\mathbf{D}} = \mathbf{D} - \mathbf{I}$, cf. derivations in theorem B.4.1 on page 149

$$\mathbf{S}_{\mathbf{xx}}^{-1} \mathbf{S}_{\mathbf{xy}} \mathbf{S}_{\mathbf{yy}}^{-1} \mathbf{S}_{\mathbf{yx}} \Phi = \Phi \tilde{\mathbf{D}} \quad (3.162)$$

The square root of the eigenvalue $\tilde{\mathbf{D}}$ are the canonical correlation coefficients.

Once the latent variables are determined the \mathbf{Q} matrix can be estimated. This can be done by a least square estimate:

$$\hat{\mathbf{Q}}_{\text{OLS}} = (\mathbf{T}^\top \mathbf{T})^{-1} \mathbf{T}^\top \mathbf{Y}. \quad (3.163)$$

If the latent variables are orthonormal the expression can be simplified

$$\hat{\mathbf{Q}}_{\text{OLS}} = \mathbf{T}^\top \mathbf{Y}. \quad (3.164)$$

The estimate is complicated by that there still might be noise in the latent variables, — especially for large K . This has led Høy, Westad, and Martens (2001) to introduce (generalized) ridge regression in the estimation of \mathbf{Q}

$$\hat{\mathbf{Q}}_{\text{GRR}} = (\mathbf{T}^\top \mathbf{T} + \mathbf{\Lambda})^{-1} \mathbf{T}^\top \mathbf{Y}, \quad (3.165)$$

where $\mathbf{\Lambda}$ is a diagonal matrix with ridge regularization hyperparameters.

Bilinear model as factor analysis

The bilinear model can be expressed as a factor analysis model (section 3.9.2), if the variances of the noise represented in the matrices \mathbf{E} and \mathbf{F} are uncorrelated. First is a matrix \mathbf{Z} constructed from \mathbf{X} with \mathbf{Y} appended

$$\mathbf{Z} = (\mathbf{X}, \mathbf{Y}) \quad (3.166)$$

Disregarding any mean for clarity, and applying equations 3.148 and 3.149, under the assumption of orthonormal latent variables $N^{-1} \mathbf{T}^\top \mathbf{T} = \mathbf{I}$ and that there is no correlation between the latent variables and the noise the covariance of \mathbf{z} is

$$\mathbf{V}[\mathbf{z}] = N^{-1} \mathbf{Z}^\top \mathbf{Z} = N^{-1} \begin{bmatrix} \mathbf{P}^\top \mathbf{P} + \mathbf{E}^\top \mathbf{E} & \mathbf{P}^\top \mathbf{Q} \\ \mathbf{Q}^\top \mathbf{P} & \mathbf{Q}^\top \mathbf{Q} + \mathbf{F}^\top \mathbf{F} \end{bmatrix} \quad (3.167)$$

Now the matrix of factor loading can be constructed as

$$\mathbf{\Lambda} = (\mathbf{P}, \mathbf{Q}) \quad (3.168)$$

and the covariance of the *specific factors* is

$$\mathbf{\Upsilon} = \begin{bmatrix} \mathbf{\Upsilon}_e & \mathbf{0} \\ \mathbf{0} & \mathbf{\Upsilon}_f \end{bmatrix}, \quad (3.169)$$

where the covariance of the noise are diagonal matrices

$$\mathbf{V}[\mathbf{e}] = \mathbf{\Upsilon}_e = \text{diag}(v_{e,11}, \dots, v_{e,PP}) \quad (3.170)$$

$$\mathbf{V}[\mathbf{f}] = \mathbf{\Upsilon}_f = \text{diag}(v_{f,11}, \dots, v_{f,QQ}) \quad (3.171)$$

$$\mathbf{C}[\mathbf{e}, \mathbf{f}] = \mathbf{0}. \quad (3.172)$$

Thus the condition on the covariance of \mathbf{z} is, cf. (Mardia et al. 1979, equation 9.2.5)

$$\mathbf{\Sigma}_{zz} = \mathbf{\Lambda}^\top \mathbf{\Lambda} + \mathbf{\Upsilon}. \quad (3.173)$$

3.18.7 Optimization of the ridge parameter and subspace dimension

A test for the subspace dimension in ordinary canonical correlation analysis can be developed (Mardia et al. 1979, section 10.2.3). For the test of no subspace $\mathbf{\Sigma}_{xy} = \mathbf{0}$ the likelihood ratio statistics has a Wilk's lambda (Andersen's U) distribution when \mathbf{x} and \mathbf{y} is Gaussian distributed

$$\Lambda(Q, N - 1 - P, P) \sim |\mathbf{I} - \mathbf{S}_{yy}^{-1} \mathbf{S}_{yx} \mathbf{S}_{xx}^{-1} \mathbf{S}_{xy}| = \prod_{k=1}^K (1 - r_k^2), \quad (3.174)$$

where r_k^2 is the square canonical correlation coefficients. Wilk's lambda is related to the mutual information between \mathbf{x} and \mathbf{y} , see derivation on page 148:

$$I(\mathbf{x}; \mathbf{y}) = -\frac{1}{2} \ln \Lambda. \quad (3.175)$$

As Wilk's lambda is not tabulated this is often replaced by χ^2 ($N \rightarrow \infty$)-asymptotic approximation.

The asymptotic approximation for the test of only S of the canonical correlations coefficients are non-zero is (Mardia, Kent, and Bibby 1979, equation 10.2.14)

$$\chi_{(P-S)(P-S)}^2 \sim - \left(N - \frac{1}{2}(P + Q + 3) \right) \ln \left[\prod_{k=S+1}^K (1 - r_k^2) \right]. \quad (3.176)$$

This approximation might not be precise when N is close to P or Q which is sometimes the case in functional neuroimaging. Other tests are proposed by (Marriott 1952) and (Gunderson and Muirhead 1997).

These tests do not directly work in conjunction with ridge regularization and singular data. Loh (1995) develops a formula for penalized discriminant analysis for the determination of what corresponds to our k_X ridge parameter. In PLS a common procedure is to estimate the dimension of the subspace with cross-validation (Wold 1978).

Bullmore et al. (1996b) and Mencl et al. (2000) used permutation test in connection with PLS to determine the significant components (subspace dimensions): The columns of the \mathbf{Y} matrix containing design variables are randomly shuffled and corresponding to a null-hypothesis and the PLS analysis is repeated, e.g., 1000 times.

Frutiger et al. (2000) used the resampling technique of (Strother et al. 1997a; Strother et al. 1997b), where the data is split and two sets of canonical correlation vectors are computed. The correlation between corresponding canonical correlations vectors are reported as a reproducibility index.

3.18.8 Generalization of canonical analysis

There exist three types of extensions to canonical analysis:

- Multiway analyses on matrices with more than two dimensions, e.g., $\mathbf{X}(N \times P \times Q \times \dots)$. Development has occurred both for unsupervised algorithms and for canonical analysis with two data sets (Bro 1998). This is of interest in connection with functional neuroimaging since data sets are often from balanced experiments, e.g., $\mathbf{X}(\text{subjects} \times \text{sessions} \times \text{runs} \times \text{conditions})$. Unsupervised analysis with a "SVD- k modes" has been applied to brain data, see, e.g., (Leibovici 2000).
- Multiset canonical analysis with more than two matrices, e.g., $\mathbf{X}(N \times P)$, $\mathbf{Y}(N \times P)$, $\mathbf{Z}(N \times Q)$. Kettenring (1971) and Sen Gupta (1982) describe multiset canonical correlation analysis.

- Nonlinear canonical analysis or “canonical manifold analysis” where the data is regarded as related through or generated from a low-dimension non-linear sub-manifold. Tibshirani (1992) describes “principal curves” which is one-dimensional sub-manifold estimated from a single set of data. A nonlinear sub-manifold of two sets of data (two matrices) can be identified with two artificial neural networks with the output in a latent space. Auto-associative models appear in (Diamantaras and Kung 1996, section 6.6). Other references are (Becker 1992; Becker and Hinton 1992; Asoh and Takechi 1994; DeMers and Cottrell 1993; Lai and Fyfe 1999).

3.18.9 Canonical analyses in functional neuroimaging

Friston, Frith, Frackowiak, and Turner (1995b) used canonical correlation analysis with $\mathbf{X}(N \times P)$ containing $N = 120$ scans represented as $P = 35$ eigensequences (from a preliminary principal component analysis of a preprocessed data matrix, P being determined as the number of eigenvalues above the mean of the non-zero eigenvalues) and $\mathbf{Y}(N \times Q)$ being a design matrix containing four sinusoid basis function (as a hemodynamic response function model) for each of three conditions: $Q = 4 \times 3 = 12$ design variables. As there are more scans than principal components and design variables the covariance matrices can be inverted. A type of χ^2 -test with correction for correlation among the scans was used to determine a significant subspace.

In the case that the task is repeated *agnostic labeling* can be used which was done by Strother, Lange, Savoy, Anderson, Sidtis, Hansen, Bandettini, O’Craven, Rezza, Rosen, and Rottenberg (1996a): The fMRI data set consisted of 8 runs each with 72 scans (24 baseline + 24 activation + 24 baseline scans). The agnostic labeling groups the first scan of the first run in the same group as first scans of the 7 other runs. Likewise with the second and following scans, so the design matrix is $\mathbf{Y}(N \times Q) = \mathbf{Y}((8 \times 72) \times (72 - 1))$. Strother et al. (1996a) also depicted the canonical variate trajectory in the first few canonical subspaces termed *state-space*. Friston, Poline, Holmes, Frith, and Frackowiak (1996b) and (Fletcher et al. 1996) also used agnostic labeling on a PET data set for canonical variate analysis with preliminary principal component analysis.

Friman, Cedefamn, Lundberg, Borga, and Knutsson (2001b), Friman, Borga, Lundberg, and Knutsson (2001a), Friman, Lundberg, Cedefamn, Borga, and Knutsson (2001c) used canonical correlation on a *local* area of the volume from fMRI: A 3×3 voxel size intraslice region is taken around each voxel. This 9 dimensional vector is then reduced by spatially symmetric filters to 5 dimensions, thus for the i ’th voxel and with 180 scans the data matrix is $\mathbf{X}_i(N \times P) = \mathbf{X}_i(180 \times 5)$. The other matrix is constructed with 6 sinusoid basis function $\mathbf{Y}(N \times Q) = \mathbf{Y}(180 \times 6)$, — this matrix is common for all voxels. As the summary image is reported the canonical correlation coefficient for each voxel and in (Friman et al. 2001a) they suggest using a test to determine a threshold.

Ledberg (1998) related two sets of volume data with a kind of canonical analysis: Splitting the data from a PET study across subjects into two new data matrices these underwent their own individual preliminary principal component analysis maintaining a number of the largest PCs. Once these are found the “closest subspace” are determined by minimizing a distance measure between the subspaces from different groups, — a method based on (Krzanowski 1979).

Bullmore, Rabe-Hesketh, Morris, Williams, Gregory, Gray, and Brammer (1996b) extracted 170 voxels from a mass-univariate analysis of 100 fMRI scans and performed singular value decomposition both directly on the $\mathbf{X}_1(N \times P) = \mathbf{X}_1(100 \times 170)$ matrix as well as on another matrix $\mathbf{X}_2(N \times P) = \mathbf{X}_2(6 \times 170)$ constructed from the regression coefficients associated with 6 sinusoidal design variables from the mass-univariate analysis. A further independent canonical variate analysis was performed on 101 voxels that were grouped according to the anatomical region they occupied, i.e., the canonical variate analysis was performed on the transposed data matrix. These voxels were represented with the 6 regression coefficients: $\mathbf{X}(N \times P) = \mathbf{X}(101 \times 6)$. The other matrix contains group indication for one of the six regions of interest (ROI) $\mathbf{Y}(N \times Q) = \mathbf{Y}(101 \times 5)$.

Frutiger, Strother, Anderson, Sidtis, Arnold, and Rottenberg (2000) used preliminary principal component analysis on 26130 voxel PET data set with 140 scans choosing heuristically the first 20 principal component for a further 10-group CVA corresponding to 10 conditions: $\mathbf{X}(140 \times 20)$ and $\mathbf{Y}(140 \times 9)$. In this study CVA was also used on behavioral data collected during the scanning.

McIntosh, Bookstein, Haxby, and Grady (1996) used partial least squares (PLS) on face processing PET data represented in the \mathbf{X} matrix and with two different design matrices as \mathbf{Y} : one with condition, the other with reaction time behavioral score. A permutation test was used to access the significance. Partial least square was also used by (Mencl et al. 2000) on fMRI scans. PLS with PET data was also investigated by McIntosh, Nyberg, Bookstein, and Tulving (1997) in a type of analysis inspired by the “seed voxel” correlation analysis of (Horwitz, Grady, Haxby, Schapiro, Rapoport, Ungerleider, and Mishkin 1992): The design matrix \mathbf{Y} is constructed from 3 representative voxels selected from 3 activated regions involved in episodic memory retrieval. Since the design consists of 4 conditions the design matrix is 12-dimensional (3 voxels \times 4 conditions). The data matrix \mathbf{X} contains the full voxel data.

Other uses of canonical analysis appear in (Busatto, Howard, Ha., Brammer, Wright, Woodruff, Simmons, Williams, David, and Bullmore 1997) (CVA on fMRI), (McCrorry and Ford 1991) (CCA on SPECT), (Lobaugh, West, and McIntosh 2001) (PLS on ERP).

Friston et al. (1996b) called the K canonical correlation vectors $\mathbf{A}(P \times K)$ *canonical images* as they were associated with P voxels. In (Frutiger et al. 2000) they are called *canonical eigenimages*. Similarly, if the K canonical correlation

N	P	Q	Type	Reference
120 scans	35 PCs	3 conditions and 6 sinusoid	SVD/CCA, fMRI	Friston et al. (1995b)
60 scans	14 PCs	12 conditions	SVD/CVA, PET	Friston et al. (1996b)
Scans	Voxels	Conditions	PLS, PET	McIntosh et al. (1996)
Scans	Voxels	Reaction time	PLS, PET	McIntosh et al. (1996)
Scans	≈ 60000 voxels	4 conditions \times 3 voxels	PLS, PET	McIntosh et al. (1997)
101 voxels	6 regression coefficients	5 ROI indicators	CCA, fMRI	Bullmore et al. (1996b)
140 scans	20 PCs	10 conditions	SVD/CVA	Frutiger et al. (2000)
180 scans	5 filter outputs	6 Sinusoid	CCA, fMRI	Friman et al. (2001b)

Table 3.18: Canonical analyses in functional neuroimaging. N , P and Q refer to the dimensions of the matrices $\mathbf{X}(N \times P)$ and $\mathbf{Y}(N \times Q)$. “PCs” is principal components. “SVD” denotes preliminary principal component analysis.

vectors $\mathbf{B}(Q \times K)$ are associated with Q time samples (e.g., fMRI) they can be called *canonical sequences*

Canonical analysis has not been used as much a univariate analysis in functional neuroimaging.

3.18.10 Example on canonical analysis

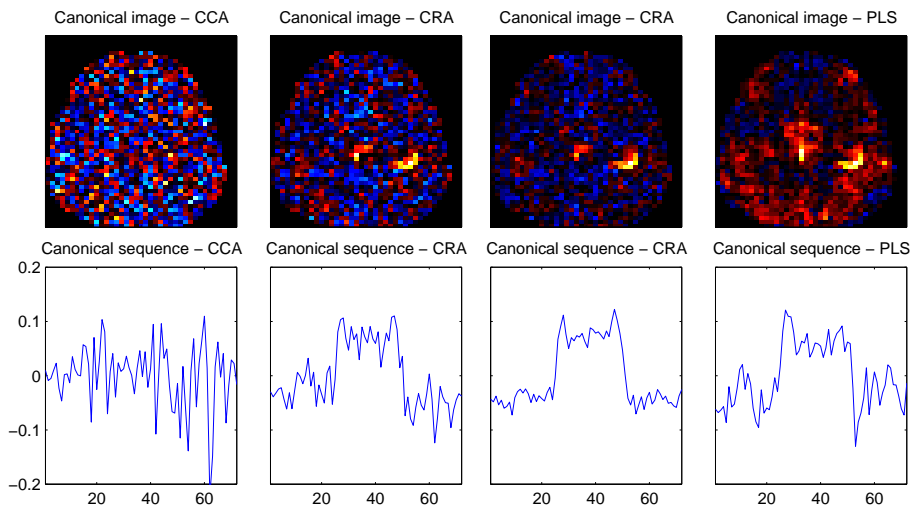


Figure 3.11: Canonical ridge analysis of an fMRI data set. The top row are the canonical images while the bottom row is the eigensequences.

Figure 3.11 shows an example of canonical ridge analysis as the ridge parameter is varied. The data is a single subject multiple runs fMRI block design where each run contains 72 scans: 24 baseline, 24 activation, 24 baseline with the activation consisting of left-handed finger-to-thumb opposition, $TR = 2.5s$ and voxel size $3.1 \times 3.1 \times 8mm$. The data was acquired on the Massachusetts General Hospital by Robert Savoy, masked and motion realigned before analysis. A $\mathbf{Y}(N \times Q)$ matrix is constructed as an indicator with data matrix with $Q = 72$ as agnostic labeling of each of the scan within the run, i.e., first scan of the first run is in the same group as the first scan of the second run. The \mathbf{X} matrix contains the fMRI image data. The top row of figure 3.11 displays the canonical images in a transversal slice through the primary motor cortex and the supplementary motor area (SMA), and the lower run the canonical sequences. The different columns represent different settings of the ridge parameter k_X for the data matrix \mathbf{X} . The left plot has $k_X = 0$ corresponding to canonical correlation analysis with preliminary SVD, and the right plot has $k_X = 1$ (c.f. equation 3.125) corresponding to the “PLS”. To two middle plot are for $0 < k_X < 1$ with the third from left being the optimal.

3.19 From image to points

When the summary image has been formed from the statistical analysis it is often summarized with a set of points or *foci*. A “foci”, “hot spot” or “location” can be defined and found in number of different ways:

- Manually, where the researcher look at the summary images and selects an appropriate voxel, e.g, the center of an activated region.

- Center of gravity (COG) of “center-of-mass”, the volume is subjected to region identification (see below) and the center of gravity (or “mass”) is found for each region. According to Posner and Raichle (1995) the COG was used by (Fox, Miezin, Allman, Van Essen, and Raichle 1987) for retinotopic mapping in a PET study.
- Local maxima, where all the voxels which immediate neighbors have a lower value are identified. This is, e.g., implemented in SPM99 (Poline 1999b): A maximum in SPM99 sense is a voxel which value is larger than any of its 18 neighbors. Only the maxima above a user specified threshold are reported and only if the maxima is far enough a part (the default is 8mm).

The center of gravity will be affected by the setting of the threshold and the type of region identification, — the maxima will not. Maxima identification can thus be made independently of the region identification, furthermore the location of the maxima is also independent of any monotonic transformation of the statistic map.

The method for separation of regions is termed *region identification*, “coloring”, “labeling” (Sonka, Hlavac, and Boyle 1993, section 6.1) or “connected component analysis” (Kirt A. Schaper). The algorithm binarizes the image from a user specified threshold and put above-threshold voxels that share the same face, edge or vertex in the same set. The different sets (regions) in the volume are labeled, e.g., with a unique integer. In 2D there are two variations of region identification: It can either be 4- or 8-neighborhood region identification, depending on whether the two voxels share the same edge (4) or the same edge or vertex (8). In 3D the variation are 6-, 18- or 26-neighborhood region identification: 6 for only face connections, 18 for both face and edge connection and 26 for face, edge and vertex connection, e.g., SPM99 implements an 18-connectivity scheme (Poline 1999b, maximum and region identification) and (Poline 1999a, only region identification). Once a region has been identified the size of it can be calculated and regions that contains a small number of voxels is sometimes discarded.

Region identification are often handled by scanning one line at a time and label regions in a first pass. While the algorithm labels the data it constructs lists of which regions are connected *between* the lines. The second pass should then merged regions across lines. Below is shown the first pass where two rows are labeled:

$$\begin{array}{cccc} 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 \end{array} \rightarrow \text{(first pass)} \rightarrow \begin{array}{cccc} 0 & 1 & 0 & 2 \\ 3 & 3 & 3 & 3 \end{array} \quad (3.177)$$

Since the regions have several different label collisions (Region 3 collide with both 1 and 2) it is not enough to keep one collision value for a region. Instead the label collisions should be kept in a list of sets. When a new label collision pair is detected either (1) a new set is constructed with the pair, (2) the pair is added to the existing set which already contain one or both of the pairs, or (3) the pairs are contained in two different sets and the two set should be merged.

A different approach to model activation foci is by modeling the statistical map, e.g., with a finite Gaussian mixture model as in (Hartvig 1999). The foci can then identified as the means of the Gaussians. It is, though, a question whether a Gaussian PDF is the right form for an activation.

A foci can be described by more than the spatial coordinate, e.g, the magnitude, the spatial extent and label. The magnitude of foci might be the regression coefficients, the test statistics, the p-value of the voxel or of the associated connected region and reported in, e.g, “ml/100g/min (absolute change rCBF)” as in BrainMap™. The extent of the activation can be described by the number of voxels above threshold in that cluster or the volume in an appropriate unit.

The foci can be labeled by the Brodmann number, the anatomical or functional area. The Talairach Atlas (Talairach and Tournoux 1988) and the Talairach Daemon (Lancaster et al. 1997c) make it possible to label the foci from Talairach coordinates, though there has been objection to this because the spatial normalization to the Talairach space is not accurate enough, see e.g., (Poldrack 2001).

Chapter 4

Visualization

Visualization is defined by Card, Mackinlay, and Shneiderman (1999, page 6) as “The use of computer-supported, interactive, visual representations of data to amplify cognition”. Branched under visualization are *scientific visualization* and *information visualization*, which are defined as:

- **Scientific visualization:** Use of interactive visual representations of *scientific data*, typically *physically based*, to amplify cognition.
- **Information visualization:** Use of interactive representations of *abstract, nonphysically based data* to amplify cognition. Among the main types of information visualization are *network visualization*, *document visualization*, *process visualization* and *information workspaces*.

Scientific visualization is used very often as a mean of reporting the result of the analysis of functional neuroimaging data. A short description of the different types are available in section 4.3. Furthermore, Robb (1999) reviews 3D biomedical visualization.

Information visualization is not used in larger extent in neuroinformatics. Young (1996) reviews 3D information visualization and Chittaro (2001) reviews information visualization in medicine. Many of the important papers in information visualization are reprinted in (Card, Mackinlay, and Shneiderman 1999).

Visualizations can be characterized by *persistence*, *distributability*, *cooperativity* and *interactivity*. If the visualization is 2D then it can be printed and “stored” on paper, thereby adding persistence to the visualization, while any interaction is lost. Persistent 3D visualization is more difficult to obtain. Usually 3D visualization is based on user interactive programs as interactive viewpoint manipulation can provide important depth information. If the visualization can be stored in a file (as the case is with VRML, see section 4.1.1) then the visualization will be persistent across viewing sessions, provided that the rendering programs still can interpret the visualization format. Distributability is closely linked to persistency: If the visualization can be stored in a file it can also be distributed, e.g., via the Internet. If the visualization is cooperative then several operators are able to interact and view the visualization.

4.1 Rendering techniques

Four main types of rendering techniques exist: 2D bitmap-based, 2D vector-based, 3D polygon based and 3D volume rendering, where the latter two will be described below.

4.1.1 3D polygon based visualization

There exists a number of programming libraries for polygon graphics, e.g., OpenGL and Direct3D. Their basic geometric element is the polygon and additionally lines and points are also supported. Further objects can be defined based on these, e.g., a mesh consisting of several polygons. The appearance of the geometry can be changed by applying texture to the surface or defining its color and transparency. Lightning can be defined with either local or global light sources, and their effect on the polygons is controlled by defining ambient, diffuse and emissive colors: For ambient color the angle between the light source and the surface normal of the polygon has no effect on the rendering, for diffuse color it has, and for emissive color the rendering depends on the angle between the viewpoint, the polygon and the light source. The shading of the polygon can be refined by Gouraud (Gouraud 1971) and Phong (Phong 1975) shading, so different areas of the polygon have different color, depending upon vertex normals, that are either defined or interpolated between connected polygons.

Realtime polygon-based graphics libraries do usually not support cast shadows, but cast shadows can be obtained by copy, scaling and translation of the geometry and provide cues for disambiguation of shapes and locations of 3D objects (Najork and Brown 1995).

Virtual Reality Modeling Language

Virtual Reality Modeling Language (VRML) is a standard for a file format that describes hyperlinked interactive 3D polygon based visualizations meant for display in *VRML-browsers*. There are actually two different standards: “VRML 1.0” (Bell, Parisi, and Pesce 1995) and “VRML97” (also called “VRML 2.0” Carey, Bell, and Marrin 1997) — the last one being an ISO-standard (ISO/IEC 14772-1:1997). The VRML file is a text file and it can contain data for description of the visible objects (geometry, its appearance, lightening, background, ...) as well as description of interaction among the objects and small programs, — with the description organized in *nodes*, — each type with its own set of properties. Apart from what is found in other standards (e.g., the Inventor format originally defined by Silicon Graphics) VRML defines hyperlinks.

VRML is *persistent* (the file is persistent but is dependent on a suitable program for visualization). VRML visualizations are also easy to *distribute*. However, if the VRML visualization contains texture, images or movie clips it is usually distributed over several files making the distribution not as convenient as PostScript and PDF files. It is not directly *cooperative* but extensions with Java can compensate for that although the persistence and distributability will decrease. VRML is partly *interactive*: The typical browser allows rotation of the world (the model) and change of viewpoint. More interaction can be obtained by using *sensors* — either interactive, e.g., plan sensors, or autonomic, e.g., the “TimeSensor” — which events can be routed to change the objects and appearance dynamically.

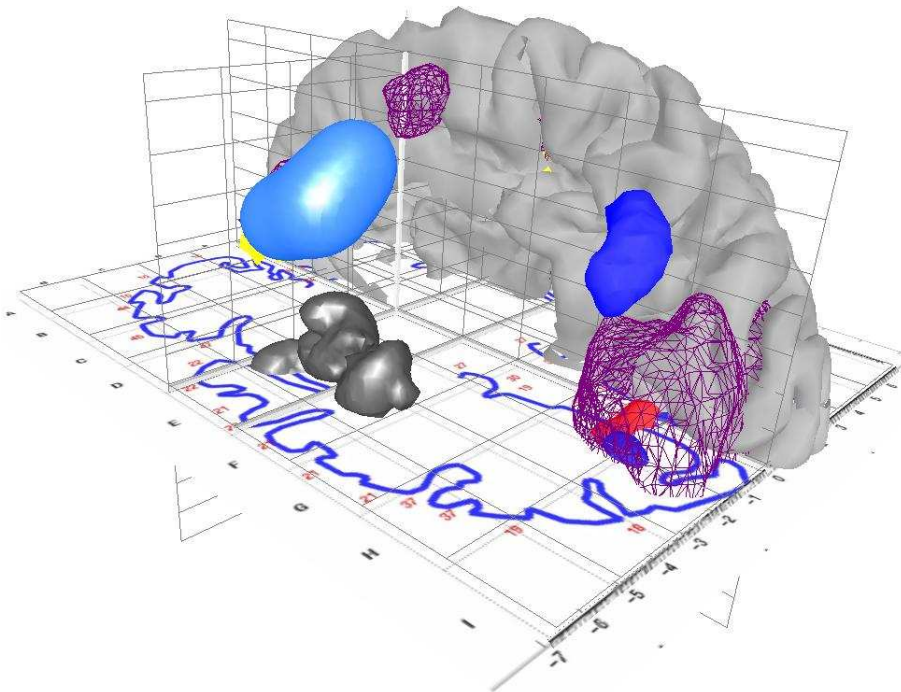


Figure 4.1: Example on a VRML visualization where different result volumes are displayed together with a digitized rendering of a page from the Talairach atlas (Talairach and Tournoux 1988) and a cortex surface constructed by Heather A. Drury, see (Drury and Van Essen 1997). The numbers on the slice from the atlas is the Brodmann numbers. This is a screenshot from the CosmoPlayer VRML-browser running on an Silicon Graphics (SGI) Onyx computer. The spatial locations of the components become obvious when the model is rotated in real-time in the VRML-browser. The result volumes are from saccadic eye movement studies (Law 1996; Law, Svarer, Rostrup, and Paulson 1998), passive word processing (Petersen, Fox, Posner, Mintun, and Raichle 1988) and olfaction (Zatorre, Jones-Gotman, Evans, and Meyer 1992, not visible) and a mouth activation study (Ian Law, personal communication).

VRML has not been used in mainstream brain mapping visualizations, but a few tools have been developed (Drury, West, and Van Essen 1997; Nielsen and Hansen 1997; Nielsen and Hansen 1998; Nielsen and Hansen 2000a), see also WebCortex (<http://zakros.ucsd.edu>) and the “NPAC 3D Visible Human Viewer” (<http://www.npac.syr.edu/projects/3Dvisiblehuman/VRML/VRML2.0/MEDVIS/>), and “3D Brain” (Kling-Petersen, Pascher, and Rydmark 1998, <http://www.mice.gu.se/MICE/education/3DB.html>). VRML has also been used as an output for an Internet-based remote Markov random field segmentation program (Kelle 1999), and for remote visualization of EEG data (Martínez, Hassainia, Mata, Medina, Leehan, and Bañuelos 2000). Figure 4.1 is an example of a VRML visualization. Others are displayed in figures 2.6, 4.3, 5.17(a), 5.17(b) and 5.18.

Text labels in the visualization are constructed using a single transparent textured polygon for a whole line of text rather than rendering each individual character as a polygon. The latter is the default most browsers use for rendering the

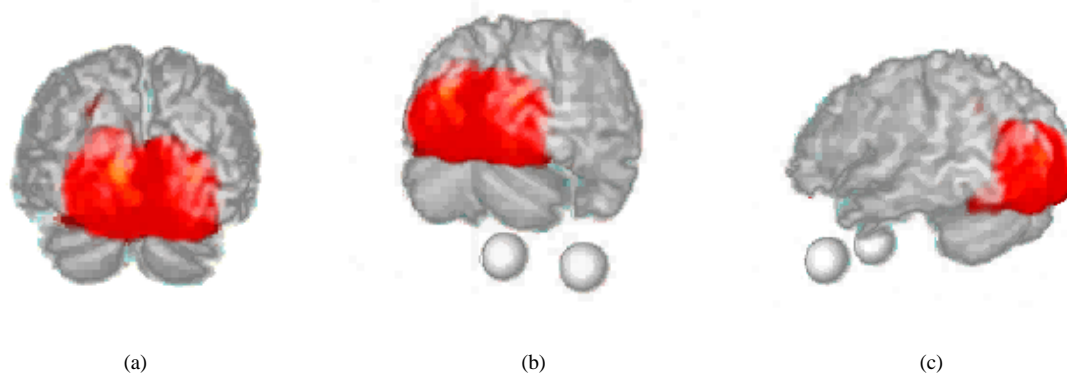


Figure 4.2: Sequence of volume renderings using the volren program with a MRI template from SPM96 and visual activation from a study of Ian Law (personal communication). The two round objects are polygon representation of the eyes.

“Text” VRML node. By the latter technique the number of rendered polygons used in the visualization is reduced and the rendering speed should increase in most browsers. The texture on the polygon are constructed from concatenated binary images of each letter. These images were constructed manually. To avoid aliasing and filter effects in the texture the size of the texture should be a power of two. In the examples the height of the letters is 8 pixels and transparent pixel columns are added to the left and right end of the constructed image so its length matches a power of two.

To orient the text in the direction of the viewer the “Billboard” VRML node is used. This node — which often is used in connection with rendering of trees in virtual reality visualization — automatically rotates its attached geometry (in this case the polygon with the text image) toward the viewer.

4.1.2 Volume rendering

Volume rendering works directly on the 3D structure. “Traditional volume rendering” is based on ray casting and is typically slow, but interactive speeds can be obtained by using hardware volume rendering. Kulick (1996) developed an interactive program (volren) where OpenGL hardware acceleration was utilized through transparent textured polygons, see figure 4.2 for an example. The technique was originally suggested by Akeley (1993) and has also been implemented by, e.g., Wilson et al. (1994) with the Voltx program. The technique produces low quality renderings as the objects are not shaded. However, new techniques have been developed which include shading, see, e.g., Van Gelder and Kim (1996), Dachille et al. (1998). Another mean for achieving interactive speeds is by dedicated volume rendering hardware (such as VolumePro™ by TeraRecon Inc., <http://www.rtviz.com>, which is supported by the Visualization Toolkit; Schroeder, Martin, and Lorensen 1997).

Tomographic reconstruction in CT and PET can be seen as the inverse operation of volume rendering and reconstruction can be based on volume rendering hardware (Cabral and Foran 1999).

Volume rendering is implemented in the AFNI program (http://afni.nimh.nih.gov/afni/plugin_render.html), see also section 4.3. Other applications of volume rendering in neuroscience appear in (Sommer, Dietz, Westermann, and Ertl 1998) that describe a system for combined texture mapping volume and polygon-based rendering suitable for medical volume visualization.

4.2 Visualization techniques

4.2.1 Polygon generation

The classical algorithm to construct a surface representation from volumetric data is the so-called Marching Cube algorithm (Lorensen and Cline 1987; Cline, Ludke, and Lorensen 1988) implemented in, e.g., polyr (Jensen 1995; Nielsen 1998). It is based on construction of “subvoxel-sized” polygons for an isosurface and produces a large number of polygons even if the complexity of the surface is low, e.g., the object contains linear planes that are larger than the individual voxels. There has been developed a number of approaches to reduce the number of polygons, see, e.g., (Jensen 1995; Shekhar, Fayyad, Yagel, and Cornhill 1996).

4.2.4 Network visualization

A simple network visualization is obtained by plotting the link matrix as an image where the color or the gray-scale of the image elements depend on the link strength. The *Hinton diagram* is a variation where the link is still arranged in a matrix but each element is rendered as a small box or circle with the size depending on the link strength.

There exists a number of ways of drawing networks when the node are represented as points and the links are represented as lines between the points, see, e.g., (Feng 1997) and figure 3.9. *Graphviz* is a useful set of programs that implement automatic 2D graph drawing (Gansner and North 2000). *dot* for directed graphs (Koutsofios and North 1996) and *neato* for undirected graphs (North 1992). See examples in figures 5.3, 5.4(a), 5.4(b) — and an application in bioinformatics: (Jenssen, Læreid, Komorowski, and Hovig 2001). Another graph drawing program is DAG (Gansner, North, and Vo 1988).

Network visualizations in connection with bibliometrics/scientometrics are available in the BIBEXCEL toolbox (Persson 2001), for examples see the author cocitation in (Persson 2000, figure 1) and the organization network in (Persson, Luukkonen, and Hälikkää 2000, figures 11 and 12). A predecessor was the BIBMAP toolbox (Persson, Stern, and Holmberg 1992). Chen (1999) describes a system for 3D VRML visualization in connection with author cocitation analysis. Both in (Chen 1999) and in (Chen 1997: 2D network visualizations of scientific documents) the *Pathfinder* network scaling method (Schvaneveldt, Durso, and Dearholt 1989) are used to reduce the network complexity, while Chen and Carr (1999) applied multidimensional scaling for author cocitation analysis. Early examples of scientometric visualizations (“SCI-map”) are available in (Small and Garfield 1985) and creating maps in connection with scientific bibliographic materials has been termed scientography (Garfield 1986).

Tree visualization

Trees represent special kind of networks. They can represent hierarchical information with no cross-links, — an example in neuroinformatics being the NeuroNames brain hierarchy (Bowden and Martin 1995).

Apart from classical visualization techniques, such as implemented in file browsers, there are a number of different visualization approaches to large scale trees, see e.g., (Card, Mackinlay, and Shneiderman 1999). The *cone tree* (Robertson, Mackinlay, and Card 1991; Robertson, Mackinlay, and Card 1994) is a 3D visualization technique that renders the children of a parent node in a circle so that the parent becomes the top of a cone and the children the base. Each level of the tree is rendered at the same *z*-coordinate. *Treemaps* is space filling maps where nodes are alternating between vertical and horizontal rows of rectangles in a 2D rendering (Shneiderman 1991; Johnson and Shneiderman 1991; Shneiderman 1992). *Hyperbolic displays* render the tree in a circle with the root node at the center and the nodes with the largest distance to the root at the periphery (Lamping and Rao 1994; Lamping, Rao, and Pirolli 1995; Lamping and Rao 1996; Lamping and Rao 1997). Inxight (<http://www.inxight.com>) terms this a *Star Tree*. Groupings can be emphasized with node or link color. The approach was also described by Munzner and Burchard (1995) who extended it to 3D rendering. *Cityscape* is a 3D visualization technique where all nodes are rendered as boxes standing on a plane and with child nodes in a circle around the parent node (Keskin and Vogelmann 1997). SimVis (Falkman 2001) incorporates several types of both 2D and 3D tree layouts used in connection with clinical data.

Sociogram — bullseye visualization

The field sociometry (Moreno 1934) within sociology and group-psychology has developed a special kind of network visualization: the so-called *target sociogram* (Northway 1940), see also (Freeman 2000) and (Sjølund 1965, section 4). The visualization has primarily been used in mapping of the social structure of a school-class. Here each pupil is requested to list whom s/he would like to work with and whom s/he would like to sit with, — and sometimes enhanced with questions like: who would you *not* like to play with. The social network from a class is then visualized in a circular diagram with a few concentric circles at different radii and with the most popular pupils positioned in the middle and the less popular in the periphery. Different symbols are usually used for boys and girls and the symbols are connected with arrows or lines depending on the pupils selections.

Wexelblat (1998) and Kato, Nakayama, and Yamane (2000) describe similar polar visualization — both in connection with Internet user path data. The *bullseye* visualization in connection with Internet search results is another similar visualization (Carrière and Kazman 1997, figure 3), see figures 5.8 and 5.9 for examples in connection with coauthor analysis. The visualization resembles the hyperbolic display though the network need not be a strict hierarchy. Related are also the TheBrain visualization technique used in WebBrain (<http://www.thebrain.com>, <http://www.webbrain.com>, Hugh 2001), Thinkmap (<http://www.thinkmap.com>, Thinkmap 2001) and VisiMap (<http://www.visimap.com>), — all which render a node at the center (first-level) child nodes in the periphery, see (Larsson and Hansson 2000) for a further description.

Node placement in node-link diagrams

The placements of the nodes form a special problem: The nodes can be placed by algorithms like multidimensional scaling operating on the link matrix, but this does not take into account that the links should also be rendered, — adding extra

terms to the optimization of the node placement:

- Avoid collinear nodes that are connected, because for some configuration it will be impossible to distinguish which points are connected. This can be resolved by using curves instead of straight lines.
- Minimize the number of crossing lines that make the visualization unnecessary complex.
- Nodes should not be placed too close together as it will be difficult to distinguish between them.

4.2.5 Stereovision

Stereovision takes advantage of the ability of the brain to reconstruct 3D information from 2D images when they are presented to and perceived differently by the left and right eyes. There exists a variety of different display systems to enable stereovision which will be described below. All of them make compromises in respect to spatial and temporal resolution, brightness, color, complexity and can be uncomfortable or inconvenient (Dumbreck and Smith 1992).

- Split-screen systems render left and right eye images next to each other either on one screen (monitor) or two, inspired by Charles Wheatstone's invention of the stereoscope in 1838. If no optics are used and the eyes are not crossed the images can only be translated a few centimeters from each other since the eyes have limited divergence ability, and this puts a restriction on the size of the image. Headsets (or head-mounted displays or helmet-mounted display, HMD) are special displays that are worn by the observer and where two small screens display images for each individual eye. A variation of this is the BOOM display (McDowall, Bolas, Pieper, Fisher, and Humphries 1990). The position of the head can be tracked, e.g., by magnetic means, and the viewpoint can be updated from this information. If the eyes are allowed to cross (so the right eye sees an image displayed on the left) then larger images can be displayed, and this has been used with 2D X-ray images (Olaf B. Paulson, personal communication).
- Anaglyph stereo uses different color filters for each eye and presents two images with different viewpoints on one screen color coded in accordance with the color filter. Usually the filters are red/cyan (red/green), and it gives a distorted color perception. Amber/blue anaglyph provides a better color perception and lower cross-talk between the stereo pairs, — especially when the color mixing is optimized (Hansen, Sørensen, and Sørensen 2000e, <http://www.colorcode3d.com>). The advantage with anaglyph stereo is that the display system requires no extra hardware, other than inexpensive glasses. An implementation issue with anaglyph stereovision systems appears as normal polygon rendering systems are not able to blend oblique polygons directly. A two-stage process is required where two images representing left and right eye viewpoints are generated in the first stage and in the second stage the color channels are merged to a single image.
- Chromo-stereography (Einhoven 1885) codes depth with the color spectrum, e.g., blue in the background, then green and finally red in the foreground. These can be viewed with special prism glasses (Steenblik 1986, <http://www.chromatek.com>). For a scientific application see (Verwichte and Galsgaard 1998).
- Polarized light display works together with polarized glasses. The filter in these glasses can either be linear or circular, with the linear filters usually arranged with the “penthouse polarization” where the left eye filters are aligned 45 degrees counterclockwise from horizontal and right eye filters 45 degrees clockwise. (Dumbreck and Smith 1992). The display system should either have two projectors or the polarization should be done by temporal switching in a filter. Full color rendition is obtained, but a considerable amount of brightness is lost (Pastoor and Wöpping 1997).
- In shutterglass stereo or “eclipse method” the display system switches between left and right eye image. This requires special hardware that synchronizes the display system with the glasses worn by the human observer. Early systems were based on mechanical switching, while now it is done by liquid crystals (Dumbreck and Smith 1992),
- Auto-stereoscopic displays are spectacles-free display systems that directly produce an image that is perceived as 3-dimensional (Pastoor and Wöpping 1997; Dumbreck and Smith 1992). This can be constructed, e.g., by using lenticular screens (Skerjanc and Pastoor 1997)

Volbracht, Domik, Shahrababaki, and Fels (1997) compare anaglyph stereo, shutter glass stereo and simple non-stereoscopic perspective viewing in organic molecule viewing tasks (e.g., count the rings in the molecule), and found a significant difference both for the viewing modality and the experience level of the viewer:

[...] viewing in the anaglyph mode shows a strong resemblance to the quality of shutter mode. A comparison of cost vs. performance of the three here discussed 3D display modes would therefore favor anaglyph stereo.

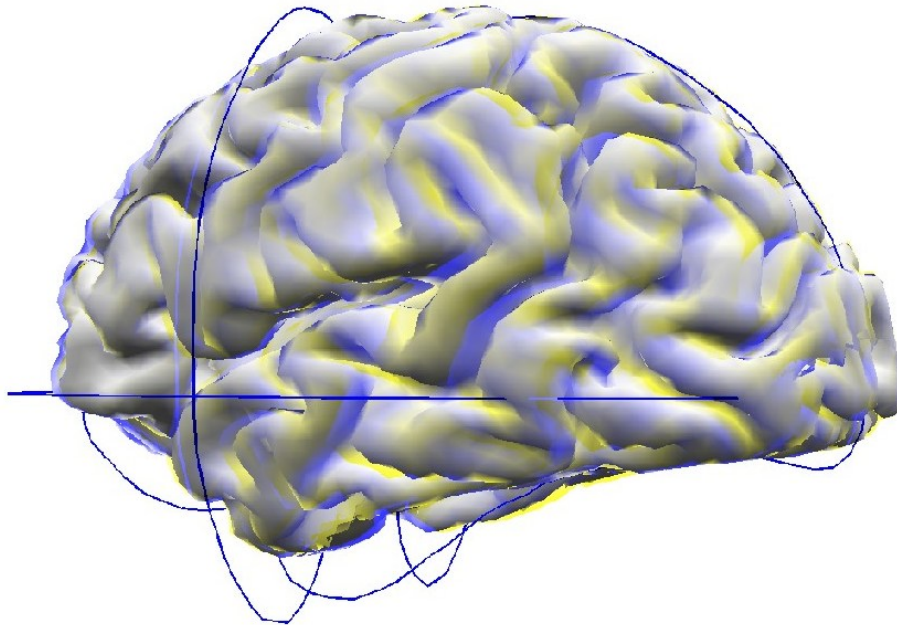


Figure 4.4: Amber/blue anaglyph stereogram showing the cortex from the segmentation in connection with the Caret program (Van Essen and Drury 1997). This will typically not reproduce well on paper. An computer-based image is presently available from <http://hendrix.imm.dtu.dk/image/>.

Headset systems in their present form are more complex to setup than simple anaglyph stereo displays. In (Nielsen and Hansen 1997) we briefly described a 3D non-stereoscopic head-track display based on modified software (CosmoPlayer, Netscape plug-in library, hardware drivers from WorldToolKit) and specialized hardware (SGI Onyx, Fastrak motion tracking, headset). Compared to ordinary 3D table top screen rendering on a fast hardware accelerated system, the HMD system had limited field of view and less quality color reproduction. Furthermore the headset was inconvenient to wear if the user should interact with other components than the visualization, e.g., another program or a piece of paper.

Figure 4.4 shows an amber/blue anaglyph stereogram constructed from Matlab. Apart from inexpensive amber/blue spectacles nothing is required to view such an image when it is presented on the computer screen. As Matlab does not directly support anaglyph images, these have to be constructed by first rendering two images corresponding to left and right eye with two different viewpoints with the built-in rendering library of Matlab. Then the two images are mixed by masking the red, green and blue components. This cannot be done real-time, but take around 20 seconds on a 700MHz Pentium computer. Multiple turnable views can be constructed in advance so that the user can get a minimum amount of interaction. Rao and Jaimes (1999) suggest a setup in connection digital photography of museum objects.

Since the cortex itself gives a good guide for the spatial location the benefit of using stereoscopy is limited if the cortex can be used as a reference. It is more appropriate when no landmarks exist, e.g., for functional activations without the support of anatomical images. MRI angiograms represents another example (Wentz, Mattle, Edelman, Kleefield, O'Reilly, Liu, and Zhao 1991).

4.3 Scientific visualization in functional brain mapping

Below is a list of the common visualization types employed in functional neuroimaging, and table 4.1 lists some of the programs for visualization.

- 2D plots. The 2D plots are usually presented in the 3 orthogonal view or as a gallery of slices (usually transversal slices).
 - Result volume alone. As the functional activation is often localized the functional volume alone present little indication of the location in the brain.

Name	Description	Reference
3D Slicer	Polygon-based visualization with alignment and segmentation	(Gering et al. 1999; Gering 1999), http://www.slicer.org
3D VIEWER	Atlas based slice viewing.	(Berks et al. 2001)
BrainMiner	3D Surface rendering and slice rendering for functional connectivity	(Mueller et al. 2000), (Zhu et al. 2001)
BrainVoyager	2D slices, 3D Polygon-based and flatmap based	http://www.BrainVoyager.de/
Caret	Flatmap	(Van Essen and Drury 1997)
Corner Cube	Surface-based rendering	(Rehm et al. 1998)
Electronic Clinical Brain Atlas	3D surface rendering and 2D slice with digitization of Talairach and Schaltenbrand atlases	(Nowinski, Bryan, and Raghavan 1997; Talairach and Tournoux 1988; Schaltenbrand and Wahren 1977)
GpetView	2D slice viewer	(Watabe 2001)
etdips	Volume and surface rendering	http://www.cc.nih.gov/cip/software/etdips/
JIV	Java-based remote 3D visualization	(Cocosco and Evans 2001)
mri3dX (BrainTools)	Slice display, flatmaps, volume rendering	http://www.mariarc.liv.ac.uk/mri3dX/
MRicro	2D Slice viewing	(Rorden and Brett 2001)
“Slice overlay”	2D slice gallery with transparency	(Brett 2000)
Spamalize	Corner Cube-like Surface rendering	http://psychz.psych.wisc.edu/~oakes/

Table 4.1: Visualization tools in brain mapping

- 2D Functional volume together with structural. This is probably the most common. The structural image is usually a MRI T1 image in gray scale while the functional volume is in color scale (often “hot”). There are different methods to merge the functional and structure volume. One way is to select a threshold for the functional volume (usually based on a p-value) and pixels from the functional volume are shown if they are above the threshold while the structural volume is shown for voxels below the threshold. Rehm, Strother, Anderson, Schaper, and Rottenberg (1994) proposed another way by using pixel interleaving where every second pixel displays the functional volume while the second half show the structural. This technique was also used as the low quality (fast) transparency technique in the VRML-browser for the SGI platform. Instead of the pixel interleaving true transparency can be used (Brett 2000). Then the transparency is not limited to 50% but can be any value between 0% and 100%. If 3D polygon graphics libraries with transparency and textures are used it is not necessary to compute the blending — this is done automatically by the library.
- 2D with structural contour. The advantage with this is that the structural image can be rendered on top of the functional image. Often used in connection with visualization of multiple activation foci from meta-analyses, e.g., (Decety and Grèzes 1999, figure 3) and (Farah and Aguirre 1999, figures 1–4).
- 2D with functional contour, e.g. (Haxby, Petit, Ungerleider, and Courtney 2000, figure 3).
- Maximum intensity projection with contour and grid. The volume is projected onto a plane with the value of the pixels on the plane determined as the maximal voxel value along the projection line. Often a threshold is applied. A black and white version with three orthogonal slices is used as the default visualization in the SPM program (Friston 1999b) and sometimes referred to as the “glass brain” visualization.

For transversal images there are two different ways to render the image (anterior part is always up): The *radiological convention* where the left brain is shown as the right side of the image, and the *neurological convention* where the right brain is shown at the right side of the image.

- Flatmap is visualization where the surface of the cortex or other brain structure is flatten to a 2D representation (Van Essen and Drury 1997). Conformal mapping can make the construction of the flatmap unique (Hurdal, Bowers, Stephenson, Sumners, Rehm, Schaper, and Rottenberg 1999).
- 3D Visualization
 - Static cortex projection is implemented in SPM99 where 3D visualization with fixed viewpoints of structural MRI are precomputed. Result volumes are thresholded and overlaid on the structural image using the same projection matrix.
 - Dynamic cortex projection can be made by projecting the result volume on a polygon model of the cortex (Herholz, Thiel, Wienhard, Pietrzyk, von Stockhausen, Karbe, Kessler, Bruckbauer, Halber, and Weiss 1996, figures 1-4) (Peter Willendrup, personal communication).

- Corner Cube environment is a polygon based visualization where the activated regions can be shown as iso-surfaces, spheres or ellipses supported by cast shadows on three orthogonal walls together with stalks rising from the “ground” transversal slice (Rehm, Lakshminarayan, Frutiger, Schaper, Sumners, Strother, Anderson, and Rottenberg 1998; Rehm, Schaper, Lakshminarayan, and Rottenberg 2000).
- Volume visualization. König, Doleisch, and Gröller (1999) describe an extension of the “Corner Cube” technique with volume rendering termed “Magic Mirrors”.

Rehm, Lakshminarayan, Frutiger, Schaper, Sumners, Strother, Anderson, and Rottenberg (1998) compared some of the visualization types in a usability study and found the Corner Cube environment to perform favorably against the gallery of slices, the maximum intensity visualization and the interleaving method (Rehm, Strother, Anderson, Schaper, and Rottenberg 1994).

The Corner Cube environment is used in several visualization, see figures 5.13, 5.15, 5.16(a), 5.16(b) and 5.20(a). These plots are constructed with Matlab where the viewpoint can be changed interactively and further interaction is added by attaching a popup menu to the geometric elements. Each component can be added or delete from the visualization either from the Matlab prompt or by clicking on the geometry. As Matlab supports downloading of Internet web-pages hyperlinking can also be made within this environment — as in VRML.

4.4 Information visualization for functional brain mapping

In many cases the information can just as well be represented as textural information rather than as an information visualization¹, and it is often easier to handle, e.g., text information is easier handled by search engines and it is easier for a user to cut and paste. However, an appropriate visual representation can enhance the “data-user” interaction, e.g., Gade (1997, pages 235–243) accounts that experiments show that it is easier to remember words that are concrete rather abstract words, presumably because these words are stored both as words and images (“double-coded”).

In (Nielsen and Hansen 1997) we described an environment termed “hyperbrain” where a functional brain mapping study is visualized, see figure 4.5 for a VRML screenshot. The visualization is based on a torus topology where stages of the study are arranged: Initial hypothesis, design of experiment, acquisition with a brain scanner, preprocessing, analysis, interpretation and thus comparison with the initial hypothesis. Among the components of the visualization are small icons representing the researchers involved in the study and with hyperlinks to their homepages. This particular model is “hand-built”, but similar models can potentially be built from information in, e.g., the BrainMapTM database.

¹Nielsen (1996b) puts the use of VRML as number 2 in his list of top ten mistakes of web design if the information does not map naturally on 3 dimensions. This might be due to users not having stable browsers or Later studies (Nielsen 1997) show that “paradigm shift” in user behavior can happen, e.g., users on the web learn scroll web-pages: Learn that web-pages can be larger than the screen

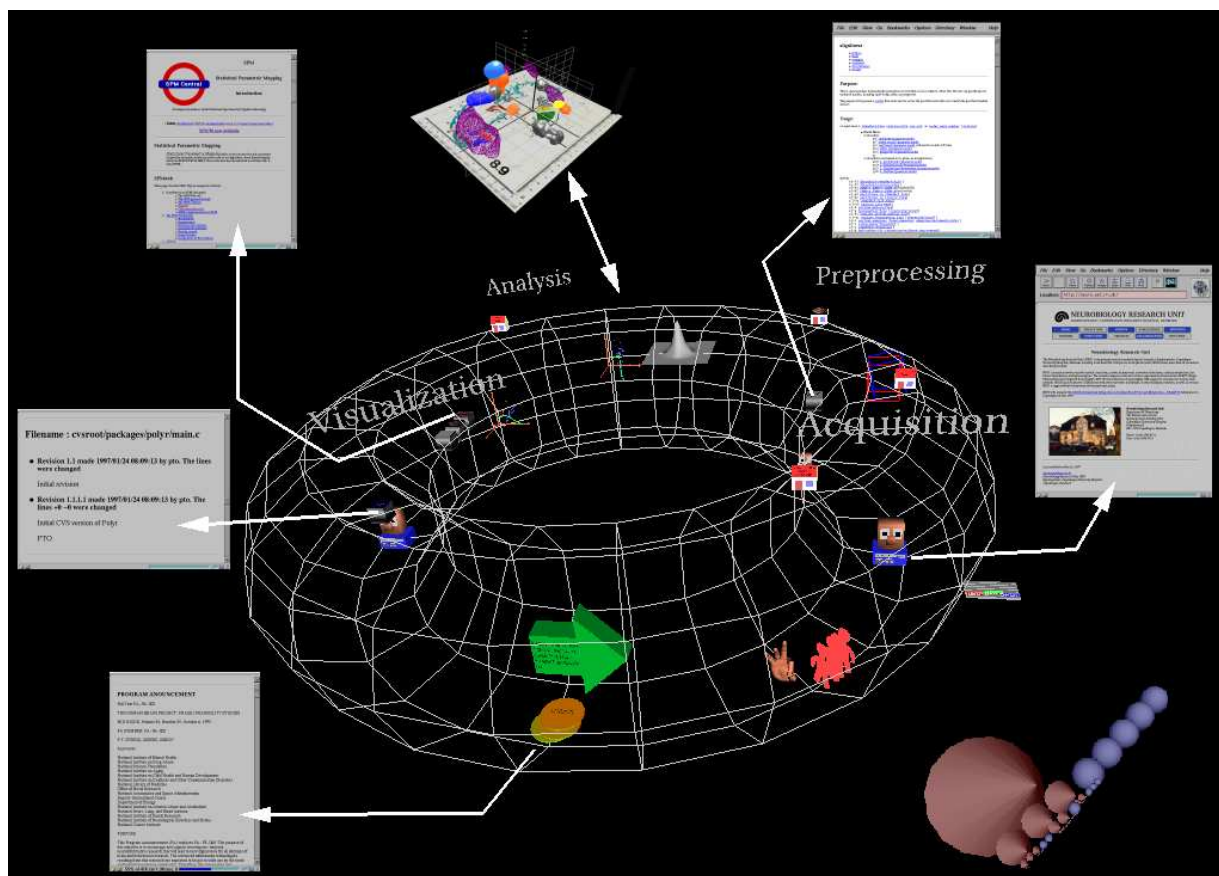


Figure 4.5: Information visualization of a functional neuroimaging experiment with linking to the homepages of the involved researchers, the used analysis packages and institution homepages.

Chapter 5

Neuroinformatics

5.1 Neuroinformatics

The term “neuroinformatics” has been used in two senses: In one sense it means the use of mathematical models inspired by neurobiology, such as artificial neural networks, in the other sense it refers to the merging of informatics and brain science with the development and utilization of computerized techniques to deal with the complex information in brain-related sciences, and it is in this second sense that I use the term neuroinformatics.

Hirsch and Koslow (1999) have defined goals for neuroinformatics:

1. Creation of new databases and query methods.
2. Development of new visualization and manipulation technologies.
3. Development of data integration and synthesis, modeling and simulation environments.
4. Development of tools for “electronic” collaboration.

Neuroinformatics can be classified as part of *bioinformatics*. This field has mainly advanced in genomics and proteomics, where DNA and protein sequences (the basic element being the four bases) are stored and analyzed (Chicurel 2000). The linear sequences data are trivial to store (e.g., in a text file) and the data are directly related to the scientific object under study. There exist several Internet-based analysis and search services in bioinformatics.

In neuroinformatics the data is much more heterogeneous, and the object under study is not directly representable (e.g., the brain area referred to as “Broca’s area” does not have a direct representation), and neuroscience produces a lot of information (e.g., fMRI and EEG measurement might only be limited by storage capacity).

Further information about the issues in neuroinformatics is available in (Chicurel 2000), (Koslow and Huerta 1997), (Shepherd et al. 1998) and a supplement edition of *NeuroImage* volume 4 number 3 December 1996.

5.2 Neuroinformatics tools for functional neuroimaging

There exists a variety of tools for functional neuroimaging tools. Perhaps the most important are the mathematical/statistical tools that are used for analysis and preprocessing of the functional neuroimages. There seems to be none of these that are web-based, but many are distributed on the Internet: These include among others SPM, AFNI (Cox 1996), AIR (Woods et al. 1998a), FSL (Smith et al. 2001), BICstat/fmristat/multistat (Worsley, Liao, Grabove, Petre, Ha, and Evans 2000), Stimulate (Strupp 1996), VoxBo (<http://www.voxbo.org/>), BrainVISA (Cointepas, Mangin, Garnero, Poline, and Benali 2001), FIASCO (Eddy, Fitzgerald, Genovese, Mockus, and Noll 1996), Lipsia (Lohmann et al. 2001), EvIdent (Pizzi et al. 2001), RPM (Aston, Worsley, and Gunn 2001), LOFA (Gokcay, Mohr, Crosson, Leonard, and Bobholz 1999), iBrain (Abbott and Jackson 2001), and Lyngby (Hansen et al. 1999b; Hansen et al. 2000b; Hansen et al. 2001b), see Gold et al. (1998) for a review.

Apart from these tools are the programs and services that deal with other stages of the study of functional neuroimages. Some of these are listed in table 5.1. Below will some of the most important be described in more detail. Apart from those listed in the table Kelle (1999) shortly reports an Internet-based Markov random field segmentation program with output to VRML, and Martínez, Hassainia, Mata, Medina, Leehan, and Bañuelos (2000) describe an Internet-based remote VRML visualization system for EEG. In appendix C on page 153 a web-service for clustering and visualization is described in more detail.

Name	Description	Reference
Brain Explorer	“Brain Atlas Database of Japanese Mon-key for WWW”. Web-based browsing of MRI	http://www.aist.go.jp/RIODB/brain/
BRAID	Image database with web-interface especially for lesions	(Bryan et al. 1995; Letovsky et al. 1998; Herskovits 2000a)
BrainMap™ ★	Web-enabled database with functional neuroimaging experiment data, e.g., Talairach coordinate	(Fox and Lancaster 1994), (Fox et al. 1995)
BrainWeb	Web-services for simulation and downloading of T1/T2/PD standard images	(Cocosco et al. 1997), (Kwan et al. 1996), (Kwan et al. 1999), (Collins et al. 1998)
CoCoMac ★	Connectivity database (for the macaque)	(Stephan et al. 2001)
Corner Cube on the Web	Visualization of Talairach locations	(Rehm et al. 2000)
Cortical neuron net database	Electrophysiology and neuron database	e.g., (Gardner et al. 2001)
ECHBD	European Computerised Human Brain Database	(Roland et al. 1999), (Fredriksson et al. 2000)
fMRIDC ★	Database with functional neuroimaging data sets and results	(Grethe et al. 2001), http://www.fmridc.org/
NeuroGenerator	Announced image database for PET and fMRI	http://www.neurogenerator.org/
NeuroInformatics Workbench	A set of neuroinformatics tools	http://www-hbp.usc.edu/workbench.html
NeuroNames	Database with structured hierarchy for neuroanatomical names	(Bowden and Martin 1995)
NeuroScholar ★	Connectivity database (for the rat brain)	(Burns 1998)
SenseLab	Web-enabled database with neuronal properties, neuronal models and olfactory receptors	http://senselab.med.yale.edu/senselab/
Talairach Daemon ★	Web-based service to map Talairach coordinates to anatomical structure and Brodmann numbers	(Lancaster et al. 2000c)
WebCortex	Interactive visualization of cortex surface	http://zakros.ucsd.edu/
The Whole Brain Atlas	CD-ROM and web-based interactive viewing of brain images	(Johnson and Becker 1997)

Table 5.1: Web-based neuroinformatics tools. The ★ indicates the tools described in detail in the text. Some of the tools are not accessible yet.

5.2.1 BrainMap™

BrainMap™ is a database with data from functional neuroimaging studies, — primarily peer-reviewed journal articles (Fox and Lancaster 1994; Lancaster, Fox, Davis, and Mikiten 1994; Fox, Lancaster, Davis, and Mikiten 1995). The database is located at the Research Imaging Center, San Antonio, Texas, <http://ric.uthscsa.edu>. Apart from the actual database there are an entry program and query services (Lancaster, Chan, Mikiten, Nguyen, and Fox 1997a). The three querying services are Internet-based with two of them implemented as ordinary programs with graphic display (Kochunov, Lancaster, and Fox 2000), and the third based on web-browsing with non-graphic CGI-queries.

Presently, the data is organized in three structures: “paper”, “experiment” and “location”, see figure 5.1. A “paper” structure encapsulates a “study” which is most often a journal article, but can also refer to an unpublished study. The structure contains bibliographic information as well as one or more “experiments”. Each “experiment” contains information about a single experiment, which in this connection is a contrast/subtraction. One of the information types is the “behavioral domain” that lists the cognitive/behavioral components believed to be involved in the experiment. The cognitive/behavioral components are ordered in a hierarchy, see table 2.1. Each “experiment” has zero or more “location”s. Each “location” is a reported hot spot representing either an activation or a deactivation.

The location web-pages each contain a single Talairach point described most importantly by the 3D coordinate (“Coordinates in Talairach, 1988 space”) that should correspond to the Talairach space (Talairach and Tournoux 1988) as much as possible, cf., the remarks in section 3.5.6. This field has been transformed from another field “Coordinates as reported in experiment”. These two field are sometimes not identical due to differences in templates. Other fields include magnitude (“Magnitude”, given in, e.g., ml/100g/min rCBF) and a p-value (“p-Value”), which in many cases appear as a

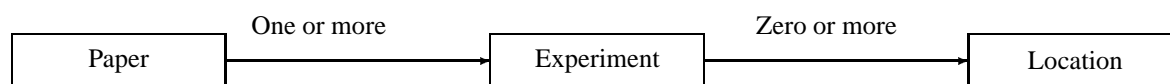


Figure 5.1: Organization of data in BrainMap™. The behavioral description is also organized as zero or more structure below each experiment, see precise description on <http://ric.uthscsa.edu/projects/brainmap.html>.

threshold value rather than the individual p-value of the location. The field “Lobar anatomy” records an anatomical label for the location as a short string, e.g., “frontal eye fields” or “frontal cortex”. Other annotation fields are “Corresponding Brodman’s area” (never set), “Functional area” (set for less than 2% of the locations, examples include “S1”, “V1”), “Lobar outline” (set for around 5% of the locations, an example is “Parietal,Lateral,AP Axis,Anterior,SI Axis,Middle”).

The “Point-Type” field “[i]ndicates whether the reported activation was determined from an individual subject’s image or from an averaged image : 1 = individual, null = mean”, the “Report/estimate” field “[i]ndicates whether a location was entered as reported or was estimated: 1 = estimate, null = reported” and the “Left/right-type” field is “[a] flag to indicate whether the reported activation falls in the left or right hemisphere of the brain: 1 = right, null = left” (Angela Uecker, personal communication, 2001-July-20).

BrainMap™ has a field in the “paper” structure to denote which “publication type” it is, e.g., “Peer Reviewed” or “Unpublished”. Some components of the “paper” in BrainMap™ is not in the original article, e.g., (Fox et al. 1996) contains no coordinates only Brodmann numbers and anatomical names, and (Petersen et al. 1988; Petersen et al. 1990) contains not all the coordinates listed in the BrainMap™ database. That is, not all data associated with a paper is directly “peer reviewed” when the “publication type” is “peer reviewed”.

A few of the papers in BrainMap™ are meta-analyses, thus it is possible that experiments in the meta-analysis papers are already recorded elsewhere in the database.

5.2.2 Talairach Daemon

The Talairach Daemon is a stand-alone program and a web-based service that resolve Talairach coordinates to anatomical names and Brodmann numbers (Lancaster et al. 1997b; Lancaster et al. 1997c; Lancaster et al. 2000b; Lancaster et al. 2000c; Lancaster et al. 2000a, <http://ric.uthscsa.edu/projects/talairachdaemon.html>). The program and service return two types of labeling: Talairach labels derived from a digitization of the Talairach atlas (Talairach and Tournoux 1988) and structure probability maps derived from a segmented brain (Collins, Holmes, Peters, and Evans 1995, ANIMAL).

The Talairach Daemon can be used to automatically label locations from functional neuroimaging statistical parametric maps. There are several problems with labeling to Talairach: 1) An exact spatial template lacks for the Talairach atlas, see section 3.5.6; 2) The individual variability in brain anatomy might not make it appropriate to base labeling on the non-probabilistic system used in the Talairach atlas being based on only a single brain.

5.2.3 Connectivity database and analysis

Connectivity databases provide information about anatomical relationships in the brain. **NeuroScholar** records connectivity in the rat brain (Burns 1998, <http://neuroscholar.usc.edu/>). Another connectivity database is **CoCoMac** (“Collation of Connectivity data on the Macaque brain”, <http://www.cocomac.org>) which records connectivity in the macaque monkey (Stephan, Kamper, Bozkurt, Burns, Young, and Kötter 2001). Such data can be “applied to interpret results from functional imaging studies” (Northoff, Richter, Gessner, Schlagenhaut, Fell, Baumgart, Kaulisch, Kotter, Stephan, Leschinger, Hagner, Bargel, Witzel, Hinrichs, Bogerts, Scheich, and Heinze 2000). None of the two database has web-interface yet.

Basically such databases consist of a $N \times N$ matrix where the elements describe the functional or anatomical connectivity (the strength or the lack of connection) between N brain regions. Further information can be incorporated in the database: The N brain regions can be organized in a hierarchy, each element in the matrix can be described by how it was obtained (bibliographic information, method, e.g., Nissl staining) and the N regions can be linked to the physical sites (brain regions), — possibly with variation information (Computational Systems Neuroscience Group (CSN), C. & O. Vogt Brain Research Institute 2001).

The data from such collections of connectivity has been used in analyses, e.g., with non-metric multidimensional scaling. For a review of this kind of analysis see (Goodhill, Simmen, and Willshaw 1995). The method allowed, e.g., Young (1992) “to comment on the segregation of dorsal and ventral processing streams and reconvergence in the DLPFC and the superior temporal area” (Friston et al. 1996a), — a hypothesis put forward by Mishkin, Ungerleider, and Macko (1983).

5.2.4 Data centers

“National fMRI data center” (fMRIDC) is a database with volumetric functional neuroimaging data (Grethe et al. 2001). The database is located at Dartmouth College, Hanover, New Hampshire and web-based with URL <http://www.fmridc.org>. The database will enable other researchers to compare results directly on the image level rather than the summary locations listed in BrainMap™. The first submission was in November 2000 by Ishai et al. (2000) and by April 12th 2001 there were 50 data sets.

fMRIDC requires authors to submit raw data (“reconstructed images from the scanner”), preprocessed data (without spatial smoothing), result images, anatomical images, and physiological/behavioral data. Furthermore, authors are encouraged to submit K-space data and raw statistical results.

Presently the small database can be searched with string queries to the bibliographic information. Presumably the search capabilities will be extended to include image retrieval techniques. Presently, some of the 50 data sets are available — not directly through download but rather by ordering and delivery with ordinary mail.

Other announced databases and data centers are the European Computerised Human Brain Database (ECHBD) (Roland et al. 1999), (Fredriksson et al. 2000) and a data center which only maintains pointers (Bly, Rebbechi, Grasso, and Hanson 2001).

5.2.5 Library-like services

There exist several services for retrieval of general textural information and bibliographical data, e.g., web-search engines such as Goggle™ enable full text search across text files on the web (HTML, PDF, ...) with ranking determined by link analysis. For a review and analysis of web search engines, see, e.g., (Lawrence and Giles 1998b) (coverage of web search engines), (Chakrabarti, Dom, Kumar, Raghavan, Rajagopalan, Tomkins, Kleinberg, and Gibson 1999) (link analysis in search engines) and (Schwartz 1998). Other types of services include directories such as the Open Directory Project (dmoz.org).

National Center for Biotechnology Information (NCBI) at the National Library of Medicine (NLM), Nation Institute of Health (NIH) provides the *Entrez* search facility (<http://www.ncbi.nlm.nih.gov/entrez/>), which incorporates a number of different bioinformatics search tools including the **Pubmed** service which interfaces to MEDLINE, — an NLM database containing bibliographic information from the biomedical literature. It is possible to construct typed queries, e.g., specifically on author names. Title, author names, citation, keywords (MeSH terms) and possibly abstract and link to the full text provided from the publisher homepage are returned together with a unique identifier (PMID).

ISI® Web of Science® (<http://wos.isiglobalnet.com/>) is a commercial Internet service that provides search in scientific references based on data from a large set of journals. The data is organized in 3 different collections *Science Citation Index Expanded™* (SCI), *Social Sciences Citation Index®* (SSCI), *Arts & Humanities Citation Index®*. Citation studies are often based on these databases.

Apart from Pubmed and ISI® Web of Science® the larger publishers often have their own search facility among their own articles together with “related document” services and unique identification through, e.g., DOI (<http://www.doi.org>).

Another effort is **ResearchIndex** — previously called CiteSeer — (Lawrence, Giles, and Bollacker 1999; Bollacker, Lawrence, and Giles 1998; Giles, Bollacker, and Lawrence 1998, <http://www.researchindex.com>), that indexes most of scientific PostScript and PDF files on the Internet within the computer science field. It is possible to do full text search within the indexed documents and “related documents” are found from link analysis, ranked by citations. Sentences are also cataloged so that documents with sentence overlap can be identified. All this is done automatically and the parsing of the PostScript/PDF files sometimes result in errors. Some of these errors can be corrected by the user through a web-interface. Other interactive components are comments and (subjective) document ranking services. ResearchIndex has little overlap in the field of functional neuroimaging, since it is not customary in the neuroscience field to publish full text articles on the Internet on public sites. Mostly methodological papers are recorded. Although PostScript is described with ASCII characters it is not directly readable, but converters exist: For a list see (Nevill-Manning, Reed, and Witten 1998). “Prescript” (Miller 1998) is fast and used in ResearchIndex but is not capable of converting certain PDF files, e.g., from Academic Press. The slower “pstotext” is able to do this (Birrell and MacJones 2000).

Citations appearing in scientific articles play a important role in assessing national research, research institution, research fields and individual researchers. Although Internet search engine results have been shown to vary (Lawrence and Giles 1998b; Snyder and Rosenbaum 1999) they have been applied in ranking of nations and organizations (Almind and Ingwersen 1997; Ingwersen 1998).

Image retrieval

Retrieval systems for other digital objects than text have also been constructed, e.g., on images. Some image retrieval systems are based on text description of images, but other have combined it with automated or semi-automated extracted features, e.g., (Djeraba and Bouet 1997) is based on color, texture, shape, spatial constraint and keywords. Web-based image retrieval systems are incorporated in, e.g., WebSEEk (Smith and Chang 1996) and AltaVista (<http://www.altavista.com>).

Liu and Dellaert (1998a, 1998b) describe image retrieval for 3D medical images specifically CT brain scans containing normal, stroke and “blood cases”. The basic “object” is a half 2D slice where features are extracted from, such as mean, standard deviation and asymmetry measures. Another medical image retrieval system is (Comaniciu, Meer, Foran, and Medl 1998). Ford, Makedon, Megalooikonomou, Shen, Steinberg, and Saykin (2001) describe briefly an “inter- and intra-study data mining” tool for fMRI activation maps. This is based on “activation signatures” to represent brain activations by size, shape, number of foci, location, orientation, and other parameters.”

Returning pages

Upon a request the usual web-services will return a web-page instantaneously. However, if the web-service takes long time to finish it is usually not appropriate to let the user wait for the web-page to be constructed. The problem with the latency can be handled in several ways:

- The result is a list and the ranking takes long time: Return *first unsorted* then sorted. This is an approach taken by the meta-search engine “Inquirus”: It first flushes the links to documents received first from the inquired search engines. When a suitable amount has been received they are ranked and flushed at the bottom of the web-page
- *Java script*. A client-side Java program communicates with the web-server program and continuously updates its content. An example of this approach is implemented in the metasearch engine “Inquirus2” (Glover et al. 1999), where links are returned continuously to a client-side Java program that show them as they arrive. As more links are received and the server has downloaded the documents and ranked the links the list of links is reordered on the client-side. A problem with this approach has been that the Java implementation in the web-browsers has not been stable.
- Set a *meta field* in the returned web-page (`meta http-equiv="Refresh"`) that requests the web-browser to update the page. The approach was taken in (Diligenti, Coetzee, Lawrence, Giles, and Gori 2000).
- *Email* when the result is finished, either the result can be returned or a link to static web-page can be returned. BrainWeb with custom MRI simulations requests use this approach.
- Return a web-page immediately with *links* to other web-pages containing the information that is delayed. This approach is taken in the simple web-service described on page 153.

If the web-service takes very long time to finish (e.g. hours or days) it is not likely that the user will continue to have the webbrowser opened on the web-page. It should be possible for the user to return to the result.

5.3 Meta-analysis

Meta-analysis is a research method where several different research studies are compared in a structured manner. U.S. National Library of Medicine defines it as “a quantitative method of combining the results of independent studies (usually drawn from the published literature) and synthesizing summaries and conclusions which may be used to evaluate therapeutic effectiveness, plan new studies, etc. It is often an overview of clinical trials.”, see, e.g., Selden (1992).

A problem with meta-analysis is publication biases, e.g., the so-called *file drawer* problem (Rosenthal 1979). Researchers (is said to) tend to only report *positive* findings, i.e., studies where the hypothesis has been confirmed. Negative findings are left in the file drawer. This effect will cause the result of the meta-analysis to be biased.

“Classic” meta-analysis is usually concerned with a one-to-one or many-to-one relation between the cause and the effect. An example of a many-to-one meta analysis is Liao (1998) where the effect on student’s achievements is investigated as an effect of using hypermedia or traditional instructions.

5.3.1 Meta-analysis in functional neuroimaging

In functional neuroimaging there is usually a many-to-many relationship to be investigated by the meta-analysis (Fox et al. 1998): E.g., the word production study of Indefrey and Levelt (1999, 2000) relates 6 “core” cognitive processes to 104 region of interests based on 58 papers, see also figure 2.2. Furthermore, the results may vary depending on the parameter choices made during the preprocessing and the analysis, thus the meta-analysis needs to be conditioned on the preprocessing and analysis methods. These issues make meta-analysis in functional neuroimaging more complex, and no principal method has been established, although some recommendations have been made by Fox et al. (1998):

- “Meta-analysis¹ should address a specific question, not troll through the data”. This is in contrast to data mining where the goal is to discover new hypotheses.

¹Fox et al. (1998) use the word metanalysis.

Domain	Papers	Exp.	Loc.	Reference
Social perception	18			(Allison et al. 2000)
Memory		10/8		(Buckner and Petersen 1996)
Cognition		275		(Cabeza and Nyberg 2000)
Visual perception of human action	6			(Decety and Grèzes 1999)
Visual recognition	19	20	84	(Farah and Aguirre 1999)
M1-area	6	6(?)	6(?)	(Fox et al. 1997b)
Episodic memory in hippocampus	52			(Lepage, Habib, and Tulving 1998)
Retrieval mode	4	5	images	(Lepage et al. 2000)
Frontal eye field	8	15	22	(Paus 1996)
Encoding and retrieval	14			(Tulving et al. 1994a)
Single word reading	11(?)	11(?)	172	(Turkeltaub et al. 2001)
Anterior cingulate cortex (cognition and emotion)	64(?)	64(?)	132	(Bush, Luu, and Posner 2000)
Middle temporal	6			(de Jong et al. 1994)
Frontal operculum (Broca's area)	9			(Fox 1995)
Anterior cingulate cortex				(Koski and Paus 2000)
Medial wall of frontal areas				(Picard and Strick 1996)

Table 5.2: Meta-analyses and reviews. Those above the horizontal line address a (set of) specific cognitive component(s) or task. Those below the line a specific brain region. The list is far from exhaustive. "Papers", "Exp" and "Loc" refer to how many papers, experiments and location was involvement in the review/meta-analysis.

- "All pertinent data should be identified".
- "Quality control should be applied judiciously".
- "Analysis methods must be established before starting the study".
- "Meta-analysis should be performed by experts in the subject matter rather than by statisticians".
- "The [included] studies should be of uniformly high quality and apply comparable methodologies". At a first level the included studies in the meta-analysis should only be from peer-reviewed journals.

Usually comparisons of results across studies in functional neuroimaging reviews/meta-analyses are done informally with tabulation and direct visualization of hot spot locations, see table 5.2 for a list of some of them and (Fox et al. 1998) for a review of reviews/meta-analyses. Apart from pure review articles ordinary scientific articles often feature informal comparison, see, e.g., (Law 1996, table 6). However, there are others that make more elaborate mathematical/statistical meta-analysis:

(Fox et al. 1997a), (Fox et al. 1997b) and (Fox et al. 1999) call their meta-analysis *functional volumes modeling* (FVM). Within this framework the location variability is modeled with a 3D Gaussian distribution on a simple activation in the primary mouth region. The size of the study (number of subject) is modeled.

Indefrey and Levelt (2000, 1999) analysed the data from 50 experiments of word production obtained with a number of different modalities: PET, fMRI, MEG cortical stimulation and subdural grid. The 50 experiments were categorized in 9 tasks (e.g., "word repetition", "word reading"), each task being decomposed into its believe set of core cognitive components, see figure 2.2. Activations were described in 28 regions at a gross level and 108 regions at a finer level. The significance of the combined activation in each region was described with a threshold on a p-value obtained through an assumption about binomial errors in each region. One of the findings were: "There also was a tendency toward fewer activations in older studies, where the technology could not reliably detect as many activations" (Indefrey and Levelt 2000, page 860). Indefrey (2001) used the same methodology on a meta-analysis of syntactic processing.

Lloyd (2000, 1999) use MDS, hierarchical cluster analysis and PCA on two data matrix containing data from the BrainMap™: One matrix $\mathbf{X}_1(N_1 \times P) = \mathbf{X}_1(36 \times \approx 200)$ and another $\mathbf{X}_2(N_2 \times P) = \mathbf{X}_2(50 \times \approx 200)$ containing data from 36 and 50 experiments where the baseline states were "eyes closed" and "looking at a fixation point", respectively. Each experiment was described by a vector with the element x_p indicating whether the experiment had activation in the p 'th anatomical structure among approximately 200 defined. The sensory modality showed up as the primary distinction and stimulus types and response contingencies as secondary.

Turkeltaub, Eden, Jones, and Zeffiro (2001) used kernel density modeling to model 172 locations from 11 PET studies with aloud single word reading. The mathematical model is essential the same as the model used in section 5.9.1. A p-value is found by computing the "volume" of the brain used as the null hypothesis. Rather than estimating this directly on the volume it is formed by a Monte Carlo technique placing random foci within the brain and estimating the probability density.

Technique	Description	References
Bibliographic coupling	Similarity between two documents as measured by the amount of citation they share	(Kessler 1963)
Cocitation analysis	Analysis of a corpus by the citation pattern.	(Small 1973)
Journal Clustering	Clustering journals, e.g., based on their citation	(Carpenter and Narin 1973; Small and Koenig 1977; Morris and McCain 1998)
Author cocitation analysis (ACA)	Analysis of citation pattern	(White 1981)
Coauthor analysis	Analysis of the coauthorship pattern	(French and Villes 1996)
Automatic indexing	Automatic construction of index	(Garfield 1972)
Automatic hyperlinking	Automatic construction of hyperlinks	(Salton 1989; Cohen 1995)
Automatic categorization	Clustering of documents	(Salton and Buckley 1989)
Stop word elimination	Ignoring frequent “non-semantic” words	(Cutting et al. 1992)
Word stemming	Removal of non-important pre- and post-fixes	(Porter 1980)
Approximate string matching	Matching of string that differ a little	(Hall and Dowling 1980)
Edit distance analysis	Measurement on how similar to strings are.	(Wagner and Fischer 1974)
Term weighting	Weighting of terms dependent on, e.g., length of document or frequency in corpus	(Salton and Buckley 1988)
Latent semantic Analysis (LSA)	PCA on bag-of-words	(Deerwester et al. 1990)
N-gram (Characters)	Analysis of fixed character sequences	(Kjell and Frieder 1992)

Table 5.3: Techniques for machine text analysis.

Herskovits et al. (1999), Vasileios Megalooikonomou and Herskovits (1999) and Herskovits (2000b) used χ^2 statistics, Fisher’s exact and Bayesian analysis of discrete variables (Buntine 1996) for data mining association between lesions and deficits in a brain-image database (BRAID), see table 5.1.

A somewhat special meta-analytic study investigated the differences in the data set between 12 different PET centers across Europe running an SPM analysis on the individual sets of data as well as pooled analysis and comparing results by z-score, graphically, with SVD and with MANOVA (Poline, Vandenberghe, Holmes, Friston, and Frackowiak 1996). Scanner sensitivity and number of subjects were found to be the most important factors of variability in the summary images.

5.4 Text analysis

There are three principal methods for machine analysis of a document (e.g., a scientific article or a web-page):

- **Content analysis** which primarily includes *term analysis* focuses on items such as the characters, words and phrases of the document.
- **Link analysis** which relates the document to other documents, e.g., through citation and hyperlinking.
- **User behavior analysis** where user patterns are analyzed. The patterns might be buying behavior or web-site navigation path which are dynamically collected, generally referred to as “navigation analysis”, see, e.g., (Kato, Nakayama, and Yamane 2000). The technique can be used to suggest related documents or tour generation (Sarukkai 2000). ResearchIndex has implemented this with the list with the heading “Users who viewed this document also viewed”.

Terms have been said to represent an endogenous and links an exogenous measure of a document and analysis of the terms alone is usually not able to assess the quality (in the sense of popularity) of a document (Kleinberg 1997). By counting the number of links to a document (e.g. scientific citations or Internet in-links) the popularity can be established. On the other hand will the number of links not be able to establish the relevance of a document.

Term and link analysis have been established in the “librarian” fields of bibliometrics, scientometrics and information retrieval mostly in connection with handling and description of scientific articles. With the advent of the Internet most of these techniques have been transferred to deal with web-pages. A number of techniques in term and link analysis are listed in table 5.3 and these will be described in more detail in the sections below.

The main application of text analysis is what is usually referred to as *information retrieval* (IR) or *information access* (IA) with Sahami (1999, section 1.1) defining three sub-applications:

- **Ad hoc retrieval** where information (documents) is returned in response to a user query. Examples of this are the Internet search engines.
- **Routing/filtering** where incoming information (documents, emails, ...) is distributed according to a users need.
- **Browsing** where the task is to order information in groups.

Other application areas are concerning the more global properties of a collection of information (which is related to browsing), e.g., ranking of authors in scientometrics.

5.5 Term analysis

There are mainly two representations in relation to computerized endogenous analysis of texts:

- **Natural language** representation (Charniak 1993; Charniak 1997). Representation is usually a parse tree. This representation will not be considered in the following.
- **Vector space** representation (Salton, Wong, and Yang 1975; Salton 1989; Sahami 1999). Representation is a vector.

In the vector space representation the information (document, or part of a document) is split up in elements called *terms*, and each term correspond to an element in a vector. The usual term element is a *word*, but also *characters* (Kjell and Frieder 1992) and *word phrases* can be terms. Word phrases and “character-phrases” (Kjell and Frieder 1992; Soboroff, Nicholas, and Kukla 1997) are often called *N-grams*, where *N* denotes the number of sub-elements (unigram, bigram, trigram, ...). If the terms are words alone the ordering of the words will not affect the vector, and the representation has been called *bag-of-words* (Mitchell 1997). The vectors of the information entities (e.g., documents) can be collected in a matrix and subjected to analysis. However, before the analysis the matrix are often preprocessed. A typical line of preprocessing from the initial textual information to the matrix to be analyzed is depicted in figure 5.2.

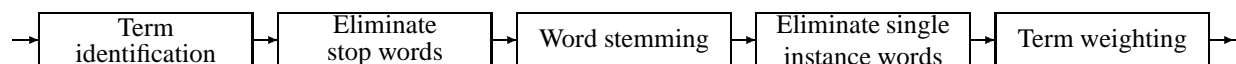


Figure 5.2: A processing scheme for the vector space representation.

5.5.1 Term identification — tokenization

Usually (defined) computer language can maintain a strict boundary between the lexical analysis and the syntactic analysis (Aho, Sethi, and Ullman 1986), and the token/term can relatively easily be found with a regular expression (While the syntactic analysis is handled by context-free parsing. More general languages in Chomsky’s grammar hierarchy — context sensitive and free — are rarely seen as computer languages). It is harder to tokenize general text such as general web-documents. The Internet search engines will usually recognize a term as a string of consecutive letters and digits. Other characters such as white-spaces and punctuations are often ignored, in the sense that consecutive white-spaces/punctuations will be regarded as a single space, e.g., searching Google™ with

"magnetic/resonance .. imaging"

will return web-pages containing “Magnetic Resonance Imaging”. Furthermore, in Internet search-engines there are typically only one type of tokens, while, e.g., Pubmed can distinguish between terms in author lists, terms in abstracts, keywords, ...

The problem of tokenization with characters other than letters and digits is complex. In cases where a certain editorial style is enforced the tokenization is relatively easy. Names (lists) and numbers are special problems. Super-pattern matching, where there is a hierarchical pattern matching (Knight 1995), will in some cases help to structure and increase the readability of the pattern matching expression. This was applied in connection with names and dates for DEADLINER with added help of dictionary look up (Kruger, Giles, Coetzee, Glover, Flake, Lawrence, and Omlin 2000, section 3.2). Very few have tried to deal with general tokenization. One of the few are (Grefenstette and Tapanainen 1994).

It is possible that machine learning techniques can be employed in learning tokens, but most systems seem to be based on hard-coded (extended) regular expressions.

5.5.2 Word stemming

Word stemming removes non-important (pre- and) suffixes from words so grammatical form has no influence, e.g., removes the 's' from verbs in English. Porter (1980) described a commonly used algorithm. However, according to Sahami (1999) Frakes (1992) found that word stemming algorithms compared to unstemmed representations in many cases perform roughly equally for retrieval tasks. A radical word stemming can be used for journal titles where each word is cut down to 3 letters, see section 5.8.3. For personal names the first and middle names can be stemmed to initials. Pubmed and ISI® Web of Science® stem to surname and two initials and non-English characters converted to English characters, e.g., “Å” to “A”. ResearchIndex does not stem and recommends searching on all variants of a name, e.g., “michael jordan or m jordan or m i jordan or michael i jordan”. For a restricted field such as functional neuroimaging the names can be abbreviated to surname and first names initial, which alleviates the problem with varying representations of authors — with and without middle name initials.

5.5.3 Stop word and single instance words elimination

Very frequent and very infrequent occurring words/terms are often discarded from the analysis, this is either for computational reasons (memory and computation time) or “statistical” reasons as they will add unnecessary complexity for the analysis, while not carrying much contextual information in the bag-of-words representation.

The very frequent occurring words are often called *stop words* and is found to be “and”, “the”, “a” etc. for the English language. A list of approximately 570 words for general texts is usually used (Sahami 1999, section 2.2.1). For specific fields the stop word list can differ, e.g., the word “significantly” appears in the stop word list of Pubmed². This word is not found in general text stop word lists.

Search engines, such as Google™, often features limited stop word elimination, and search queries like “to be or not to be” will not return anything relevant to the query.

Often infrequent words/terms are also discarded from an analysis. Using Zipf’s law (Zipf 1949) it can be shown that terms that appear only once in the corpus can account for approximately half of the unique terms (Sahami 1999, section 2.2.2). Zipf’s law states that the product of the relative frequency f_q of a term q and its rank r_q is approximately a constant c

$$c \approx r_q f_q, \quad (5.1)$$

The constant c has been found to be around 0.1 for different corpus, see, e.g., (Croft 2001).

5.5.4 Term weighting

When a bag-of-words (or bag-of-terms) vector is built from the text each element usually contains either the binary identifier indicating if the term appear in the text or the frequency by which it appears. The frequency-based vectors can be weighted by various functions: Table 5.4 shows some weighting functions described by Sahami (1999, section 2.1.2) and (Salton and Buckley 1988, table 1).

Perhaps the most well-known is *tfidf* where *tf* and *idf* are combined (Salton and Yang 1973).

$$x_{qn} = f_{qn} \cdot \log(N/N_q). \quad (5.2)$$

In (Salton and Buckley 1988) this is shown to perform well when combined with cosine the normalization.

$$\mathbf{y}_n = \mathbf{x}_n / \sqrt{\mathbf{x}_n^T \mathbf{x}_n}. \quad (5.3)$$

5.5.5 Other document elements

Apart from using the terms in the document the **formatting** can also be used. Some web search engines give more weight to terms if they appear in the <title> or large field in the HTML document. This is prone to spamming, as sites that wants to get high rating can edit local documents to conform the web-search engine ranking method, e.g., by using invisible text (Chakrabarti et al. 1999).

Lix is a simple heuristic number for the readability of a text devised by Björnsson (1971). It is defined as the sum of the percentage of words longer than seven letters and the mean number of words in each sentence. A closely related number (termed “gunning fog index”) is devised by Armstrong (1982):

$$g = 0.4 * (s + w), \quad (5.4)$$

where s is the average sentence length and w is the percentage of words with three or more syllables (not including prefixes or suffixes).

²The PubMed stop word list is available at <http://www.ncbi.nlm.nih.gov/entrez/query/static/help/pmhep.html#Stopwords>.

Functions	Description	Reference
Term frequency		
$a_{qn} = 1$	Binary, indicating if the term is present	
$a_{qn} = f_{qn}$	Raw frequency — term frequency (tf)	Luhn (1957)
$a_{qn} = 0.5 + 0.5 f_{qn} / \max_q (f_{qn})$		
$a_{qn} = \log(f_{qn} + 1)$		Robertson and Sparck Jones (1976)
$a_{qn} = \sqrt{f_{qn}}$		Cutting et al. (1992)
$a_{qn} = f_{qn} / (f_{qn} + \text{constant})$		Robertson and Walker (1994)
Collection frequency component		
$b_q = 1$		
$b_q = \log(N/N_q)$	Inverse document frequency (idf)	Sparck Jones (1972)
$b_q = \log[(N - N_q)/N_q]$	“Probabilistic inverse collection frequency”	
Normalization		
$c_n = 1$	No normalization	
$c_n = 1/\sqrt{\mathbf{x}_n^T \mathbf{x}_n}$	Cosine normalization	

Table 5.4: Weighting functions for frequency-based vectors. f_{qn} is the (absolute) frequency of q 'th term in the n 'th document, N is the total number of documents and N_q is the number of document where the q 'th term appear. Entries taken from (Sahami 1999, section 2.1.2) and (Salton and Buckley 1988, table 1).

5.5.6 Analysis techniques for term matrices

Normal information retrieval with vectorized representations of documents will compute the inner product between vectors weighted by some of the functions listed in table 5.4: One vector represents a query term and the other is taken from a matrix in the corpus to be searched. This is, e.g., implemented in Pubmed (Wilbur and Yang 1996, <http://www.ncbi.nlm.nih.gov:80/entrez/query/static/computation.html>).

“Latent semantic-analysis” (LSA) or “latent semantic indexing” (LSI) perform singular value decomposition on the documents-by-term matrix consisting of vectorized representation of documents (Deerwester, Dumais, Furnas, Landauer, and Harshman 1990; Deerwester, Dumais, Furnas, Harshman, Landauer, Lochbaum, and Streeter 1989). When a subspace is used it is possible to partially reduce the problem with synonyms, e.g., queries on “fMRI” will find documents which mentions “functional magnetic resonance imaging” and not necessarily “fMRI”. LSA can also be used to visualize the structure of the corpus. Furthermore, some of the principal components might correspond to subgroups of documents, i.e., documents with a similar content. However, the orthogonality constraint of SVD might not make it optimal for this. Application of independent component analysis (ICA) on the document-by-term matrix has also been described (Isbell Jr. and Viola 1999; Kolenda and Hansen 1999).

Apart from these techniques many other multivariate techniques can be applied: Cluster analysis or non-negative matrix factorization and — if document classes is known for a training set — linear and non-linear regression and classification methods

5.6 Link analysis

Link analysis (or link-structure analysis) is analysis of “connection” between “objects”. Here the connection will be called *links* and the objects for *nodes*. In graph theory they are usually called *vertices* and *edges* (or arcs), respectively, and in percolation theory links are called bonds. Whereas Bayesian belief networks or so-called graphical models are concerned with establishing a hidden and unobserved network structure from a set of data, link analysis starts from the network data and the goal is to infer some characteristics of its structure.

The links can either be directed or undirected. In the first case the network represents a directed graph or *digraph*, and one can distinguish between in-links (sink) and out-links (source). Bibliometrics/scientometrics distinguish between analysis on the in-links (cited documents) called *cocitation analysis* (Small 1973) and analysis of out-links (citing documents) called *bibliographic coupling* (Kessler 1963). However, in many cases this distinction is not necessary as the same mathematical devices are used in the analysis.

The network can be represented as either an adjacency matrix (sometimes called a “topology matrix”), a data matrix, a distance or similarity matrix. If the network is represented in a quadratic adjacency matrix $\mathbf{A}(N \times N)$, then the rows and columns refer to N nodes. In the case with a directed graph the rows will represent out-links and the columns in-links. If the nodes consist of two groups — one set of nodes only have out-links and no in-links, the other only in-links and no

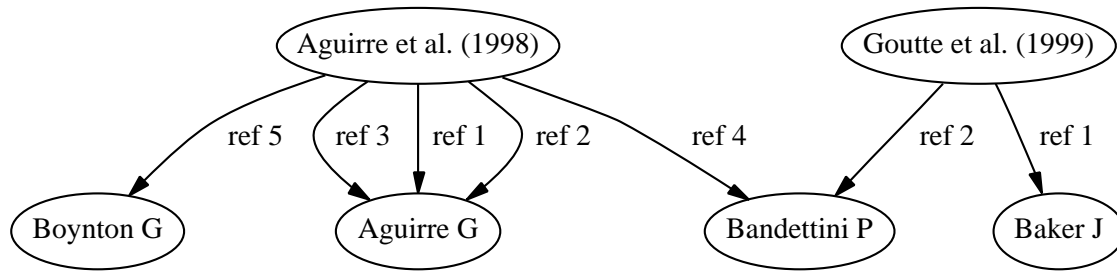


Figure 5.3: Graph in connection with author cocitation analysis, see section 5.6.3 and figure 5.5. A few of the first out link (citations) from (Aguirre et al. 1998b) and (Goutte et al. 1999b) to four first authors, with both citing documents and cited authors as nodes and the references as the links. Constructed with the *dot* program, see section 4.2.4.

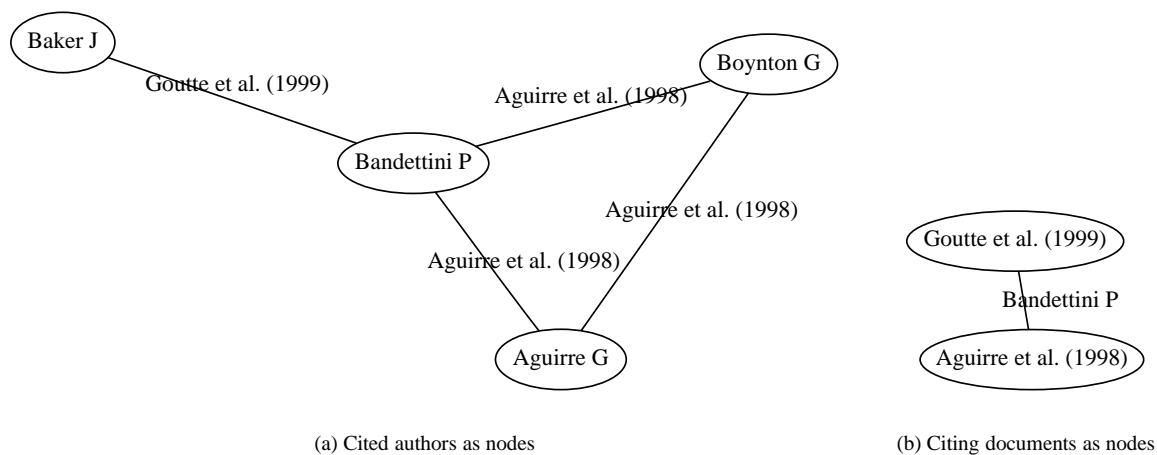


Figure 5.4: Cited authors and cited documents as nodes: see section 5.6.3 and figure 5.5. A few of the first out-links (citations) from (Aguirre et al. 1998b) and (Goutte et al. 1999b) to four first authors, with both citing documents and cited authors as nodes and the references as the links. Constructed with the *neato* program, see section 4.2.4.

out-links — the adjacency matrix is a block matrix, that can be collapsed into a data matrix: Consider author cocitation analysis (see below) where a set of M documents (scientific articles) cites Q authors. If the data is represented in an adjacency matrix $\mathbf{A}(N \times N)$ the size is $N = QM$, see figure 5.3 for a graph equivalent to this representation. But as the documents cites authors and not the other way round the \mathbf{A} matrix will only have values different from zero in one of the four quadrants and this submatrix can be extracted to an (in most cases) non-quadratic data matrix $\mathbf{X}(M \times Q)$. The data matrix and adjacency matrix can either be an *indicator matrix* with ones in the element x_{ij} when the j 'th author being cited by the i 'th document, or it can be an *abundance matrix* (Mardia et al. 1979, exercise 13.4.5) with the counts of the citations the j 'th author gets from the i 'th document.

In connection with a similarity matrix $\mathbf{C}(N \times N)$ the element c_{ij} can be set to the number of documents that cites the i 'th and j 'th author. It is always possible to construct a distance matrix from a data matrix, e.g., by a simple matrix product $\mathbf{C} = \mathbf{X}\mathbf{X}^T$.

It is not always obvious how the data should be oriented in the data matrix: In the above example with author cocitation the authors can be considered the N objects while citing documents are the P dimensions, but it might as well have been the other way around if the relationship between the documents was of interest, see figures 5.4(a) and 5.4(b).

There are several different analyses methods in link analysis. As the link data can be represented as an ordinary data matrix many of the statistical methods described in chapter 3 can be used, e.g., singular value decomposition. Other methods that are specifically designed for link analysis originates in stochastic processes with Markov chains (Cox and Miller 1965), network theory, graph theory, “graph algorithms” (Kingston 1990, chapter 10) and percolation theory, see, e.g., (Snyder and Steele 1995). Of interesting methods here are the following:

- **Ranking** aims at describing important nodes in a network. Present techniques involve simple counting and eigenvalue decompositions, e.g., in connection with the *hubs* and *authorities* method (see below). This is related to *seriation* for “ordinary” data (Mardia et al. 1979).

- **Categorization/clustering** aims to find subgraphs, connected components or communities, e.g., Gibson, Kleinberg, and Raghavan (1998): “web-communities”.
- It is often interesting to start the analysis *from a specific node*, e.g., “related documents” as implemented in ResearchIndex is based on link analysis starting from the individual node (document). Within mathematical folklore the “Erdős number” considers the co-authoring graph spanned from the famous mathematician Paul Erdős: With authors as vertices and documents as edges the Erdős number is the number of coauthored documents between an author and Erdős.
- **Global aspects** of the graph can also be considered, e.g., the connectedness, percolation and *diameter*. The diameter is defined as the longest simple path in the graph. If the whole graph is not one connected component then the diameter is infinite. Alternatively, each of the connected components can be characterized with a diameter. An example: With data from a crawl in 1999 Broder, Kumar, Maghoul, Raghavan, Rajagopalan, Stata, Tomkins, and Wiener (2000) found that a path between two web-pages existed for only 24% of them.

If the network is *dynamical* this can also be characterized, e.g., in scientometrics “immediacy” is a measure for how quickly the average article gets cited.

Categorization/clustering can be combined with ranking, e.g., by first identifying the nodes in a subgroup and after that ranking the nodes within the group. These two methods will be described in more detail in subsections below.

The links in the graph can be *typed*, e.g., scientific citations can have reasons for citation: “giving credited for related work”, “identifying methodology, equipment, and so on”, “criticizing previous work”, “disclaiming work or ideas of others” (Nanba, Kando, and Okumura 2000). Internet web-page may furthermore have navigational hyperlinks. If the type of the link can be found, e.g., by analyzing terms around the hyperlink as in (Nanba, Kando, and Okumura 2000), the analysis of the link structure can be further enhanced.

Link analysis can, e.g., be used to analyse telephone calls, social networks or computer networks. Among the applications that are of interest in connection with neuroinformatics are:

- **Cocitation analysis** can find related documents by cocitation analysis. Both the in- and out-links can be used (Pitkow and Pirolli 2001), and they can, furthermore, be combined (Bichteler and Eaton III 1980). Cocitation analysis is, e.g., implemented in the ResearchIndex database.
- **Author cocitation analysis** can be used to identify communities in a scientific field (White 1981), and in scientometrics to evaluate authors. This will be described in more detail in section 5.6.3.
- **Coauthor analysis**, see e.g. (French and Villes 1996), can give an overview of the collaboration between a set of authors.
- **Journal cocitation analysis** (Morris and McCain 1998) can rank and cluster journals.
- **Internet hyperlink analysis** can rank pages for search engines.
- **Analysis of neural pathways** from connectivity databases as those listed in section 5.2.3.

Analysis of metabolic networks where substrates are nodes and chemical reactions are links (Bilke and Peterson 2001, <http://wit.mcs.anl.gov/WIT2/>) could also be of interest for neuroscience. Another interesting application is cocitation analysis of genes (Jenssen, Læreid, Komorowski, and Hovig 2001; Pearson 2001) to resolve nomenclature differences among gene names. In this particular study 13,712 human genes and 10,125,978 MEDLINE records founded the basis for the analysis with a web-service available at <http://www.pubgene.org/>.

5.6.1 Ranking

Authorities and *hubs* are “key” nodes in the network, the terms being introduced by Kleinberg (1997, 1998, 1999). Authorities are nodes that receive many in-links and hubs are nodes that have many out-links. A simple definition might just count the number of in- and out-links. A refinement normalizes by the number of citable items, see, e.g., (Ingwersen 1998) for calculation of web impact factors. However, in the algorithm devised by Kleinberg (1997) the hubs and authorities are mutually reinforcing, so that links received from (strong) hubs and out-links from (strong) authorities weigh more. The result is the principal components of a singular value decomposition of the adjacency matrix \mathbf{A}

$$\mathbf{ULV}^T = \text{svd}[\mathbf{A}] \quad (5.5)$$

If the out-links are in the N rows and the in-links are in the P columns of a matrix \mathbf{A} then the authorities will be the elements with the highest score in the first right principal component \mathbf{v}_1 and the hubs are the elements with the highest score in the left principal component \mathbf{u}_1 (Kleinberg 1997, theorems 3.1 and 3.2), (Mardia et al. 1979, theorems A.6.2 and A.6.5). Authorities and hubs are returned separately from the Inquirus meta-search engine (Lawrence and Giles 1998a).

Another decomposition technique is used in the Goggle™ Internet search engine (<http://www.google.com>) with an algorithm termed PageRank™ (Page 1997; Brin and Page 1998). The graph is considered a Markov chain (Cox and Miller 1965, chapter 3) by transforming the adjacency matrix to a *stochastic matrix* by making the row sums to unity:

$$\mathbf{B} = \mathbf{A} \oslash (\mathbf{A} \mathbf{1} \mathbf{1}^T), \quad (5.6)$$

where \oslash denotes element-wise division. The Markov chain corresponding to this stochastic matrix \mathbf{B} will typically have absorbing states (corresponding to web-pages with no out-links). These states will dominate the solution, even though they have few in-links. To cope with this the actual PageRank™ algorithm is augmented corresponding to that all web-pages have a small out-link to all other web-pages. This will produce the desired result that the web-pages with many authoritative in-links dominate the solution.

After Kleinberg's and Goggle™'s suggestions other ranking schemes were developed, see, e.g., (Henzinger 2000; Borodin, Roberts, Rosenthal, and Tsaparas 2001).

5.6.2 Categorization and clustering

It is of interest to find subgroups of nodes. This can be done either supervised (as categorization) or unsupervised (as clustering). The decomposition, factorization and cluster analysis algorithms from chapter 3 can be used for clustering, and, e.g., neural networks can provide methods for supervised analysis. In extension to these, algorithms that are focused on the data as a graph have been suggested:

- Botafogo and Scheiderman (1991) consider identification of the so-called *articulation points* in a graph. These are special key node that form a bottleneck between two clusters, see also (Pitkow 1997).
- The *minimum cut* algorithm (Flake, Lawrence, and Giles 2000).
- Clustering can be performed by minimizing the graph diameter of the subgroups, see (Doddi, Marathe, Ravi, Taylor, and Widmayer 2000).
- If the subgroups are complete separated algorithm like that of *region identification* (connected components) can be used, see section 3.19. Usually graphs such as spanned by Internet hyperlinking, scientific citations and coauthoring patterns show a high degree of connectedness, dominated by one single component. However, for a limited subset there might be suitable separated components, see, e.g., section 5.8.2.

5.6.3 Author cocitation analysis

Author cocitation analysis (ACA) is the analysis of how two different authors are cited together. It is useful for getting an overview of a (scientific) field, and for identifying key authors (e.g., for reviewing) and “to give meaning to such abstract words as ‘influence’, ‘impact’, ‘centrality’, ‘marginality’, ‘speciality’, ‘interdisciplinarity’, ‘paradigm’, ‘shift’ ” (Lunin and White 1990). ACA was introduced by Howard D. White (White 1981; White and Griffith 1981), — for a short history see (Lunin and White 1990). The *Journal of American Society for Information Science* devoted a part of volume 41 issue 6 in 1990 to author cocitation analysis with an introductory paper by McCain (1990).

ACA is essential link analysis and common tools have been non-metric multidimensional scaling, hierarchical clustering and factor analysis. It has been applied to fields such as macro-economics (McCain 1990), information science (White and McCain 1998), neural network research (McCain 1998), family and marriage sociology (Bayer, Smart, and McLaughlin 1990) and decision support systems (Eom and Farris 1996). Apart from the static structure of the field ACA can also characterize dynamical aspects, e.g., McCain (1998) found an increasing separation between natural sciences/psychology and engineering/neural networks research in the neural network research field under study. A related field is *journal cocitation analysis* (JCA) where the focus is on cited journals rather than cited authors (Carpenter and Narin 1973; Small and Koenig 1977; Morris and McCain 1998). Cocitation analysis has also been applied to web-pages (Larson 1996).

McCain (1990) accounts the different stages in author cocitation analysis that may well apply to other types of cocitation analysis. Here will be regarded the following steps:

- **Selection** of documents, authors, website. Usually the number of items selected is restricted, e.g., McCain (1990) restricts her analysis of authors within macro-economics to 41 leader researchers, Bayer, Smart, and McLaughlin (1990) considered 36 authors in the area of marriage and family sociology. In connection web-page cocitation analysis did Larson (1996) examine 34 web-pages in the area of earth science, remote sensing and geographic information systems. For journal cocitation analysis Morris and McCain (1998) examined 29 journals in the field of medical informatics. Table 5.5 lists some of the key authors in functional neuroimaging. The figures are the values obtained from ISI® Web of Science®. The authors were selected manually by choosing the authors with a large number of abstracts in the proceedings from the International conference of Human Brain Mapping (Belliveau, Fox, Kennedy, Rosen, and Ungeleider 1996; Friberg, Gjedde, Holm, Lassen, and Nowak 1997; Paus, Gjedde, and Evans 1998).

Author	Citations	Author	Citations	Author	Citations
Alpert NM	928	Fukuyama H	(2884)	Rao SM	1442
Anderson JR	?	Gjedde A	1615	Reutens DC	214
Andreasen NC	3293	Gore JC	1159	Renshaw PF	337
Andermann F	1912	Grood W	307	Roland PE	1511
Ashburner J	132	Hammeke TA	484	Rosen BR	2944
Bandettini PA	866	Hennig J	(1252)	Rottenberg DA	535
Beaudoin G	(1177)	Herzog H	(1242)	Sadikot AF	253
Belliveau JW	1485	Hillyard SA	980	Seitz RJ	696
Bookheimer SY	244	Holmes AP	124	Sereno MI	379
Bosch J	2334	Kawashima R	304	Sharma T	(370)
Brammer MJ	285	Kennedy DN	1298	Simmons A	(1334)
Buchel C	122	Lancaster JL	318	Strother SC	329
Buckner RL	317	Lange N	293	Taylor JG	385
Bullmore ET	61	LeBihan D	1284	Thompson PM	423
Caviness VS	1360	Mangun GR	356	Toga AW	396
Chugani HT	880	McIntosh AR	580	Turner R	3471
Cohen MS	(3005)	Mazoyer BM	371	Tzourio N	343
Cox RW	172	Mazziotta JC	2799	Ugurbil	1720
Crawley AP	209	Miller MI	434	Ungerleider LG	1830
Dale AM	290	Muellergaertner HW	68	VanEssen DC	1507
Dolan RJ	1625	Neville HJ	387	Vanhorn JD	87
Drury HA	44	Noll DC	336	vonCramon DY	186
Duyn JH	314	Paulson OB	1202	Weiller C	854
Evans AC	2347	Paus T	261	Weinberger DR	3423
Fischer H	1114	Petersen SE	2373	Weisskoff RM	1611
Fox PT	2000	Petrides M	1065	Williams SCR	308
Frackowiak RSJ	5318	Pike GB	233	Woldorff MG	163
Frank JA	1539	Poline JB	272	Woods RP	878
Freund HJ	(2154)	Posse S	257	Worsley KJ	483
Friston KJ	2977	Price CJ	454	Zilles K	1684
Frith CD	3266	Pugh KR	157		
Fukuda H	1824	Raichle ME	2647		

Table 5.5: Author citations in ISI® Web of Science®. Collected 1998-Oct-30 and 1998-Nov-2. Numbers in parenthesis probably contain citation numbers for more than one author. This is due to several author sharing the same ISI abbreviated name. There seems to be name conflict at least in the cases for AC Evans, BR Rosen, JG Taylor. For JR Anderson, G Beaudoin and perhaps MS Cohen, e.g., there are severe name conflicts, i.e. where most of the articles (probably) refer to another author. It is time consuming to resolve these conflicts. Manual or automatic procedures would examine the title and the name of the journal (and the perhaps the terms in the abstract) and cluster the documents. In some case writing to the author might be the only way to resolve the conflict.

- **Retrieval.** The construction of the cocitation data for authors and journal article can be costly. The classical ACA — as described by McCain (1990) — relies on data from the Institute for Scientific Information (ISI) with the *Science Citation Index* (which include the web-service ISI® Web of Science®). The queries have to be performed with specialized costly services, thus restricting the number of authors that can be obtained and the number of analyses that is possible to do. The data will usually be a similarity matrix with the number of times two authors are cited in the same article.

Now other means are available to obtain cocitation data, e.g., from the ResearchIndex database (Lawrence et al. 1999). With this data it is possible to get the author \times document data matrix. Many journals are now available from the publishers homepage. Unfortunately, it is usually prohibited to systematically download the full text. Some publishers, e.g. Academic Press in Idealibrary (<http://www.idealibrary.com>), separate the references from the full text and are probably less restrictive about downloads of these pages, see section 5.8. Journals where the author pays the publishing cost — such as Proceedings of the National Academy of Sciences (<http://www.pnas.org>) — are less restrictive.

- **Metric.** Before the analysis of the cocitation data the data can be scaled. (McCain 1990) mentions that the diagonal element of the cocitation matrix (the similarity matrix) might be orders of magnitude larger than off-diagonal elements, and a method “adjusted diagonal values downward by substituting a value based on the highest off-diagonal cocitation counts for each author”. When the original data with citing documents exists other means of scaling the data can be done, e.g., down-weighting documents that contain many citations.

- **Analysis** by principal component analysis, factor analysis, cluster analysis and/or multidimensional scaling. These methods are the same as described in sections 3.8 and 3.9.

5.7 Combining link and term analysis

Link and term analysis can be combined: “HyPursuit” uses both content and link information for clustering hypertext (Weiss, Velez, Sheldon, Nemprempre, Szilagy, Duda, and Gifford 1996), and WebQuery also uses terms and links for visualization (Carrière and Kazman 1997). Kleinberg (1998) started with simple term analysis provided by the result from a query to a web search engine. The set of the highest ranked web-pages returned (typically 200) termed the *root set* was augmented by the web-pages with out- and in-links to the web-pages in the root set; the augmented set termed the *base set*. In the Goggle™ search engine queries can be made with terms and the documents are ranked according to a link analysis. Rafiei and Mendelzon (2000) and Mendelzon and Rafiei (2000) used a random walk on web-pages similar to the one in Kleinberg’s algorithm but conditioned the jump between nodes of the graph on whether the web-page contained a specific term. By this method it is possible to determine what topic a web-page is known for.

Cohn and Hofman (2001) performed term-based “probabilistic latent semantic analysis” (PLSA, Hofmann 1999) which is closely related to NMF (section 3.10) on the document-term matrix together with link analysis based on PHITS (Cohn and Chang 2000). A parameter in this model controls on a continuum how much a factorization depends on the terms versus the links.

The *generalised similarity analysis* (GSA) of Chen (1997) uses both terms (content), linkage and usage information in structuring and visualizing HTML documents.

5.7.1 Context sensitive author ranking

Authors can be ranked conditioned on one or several terms, here termed *context sensitive author ranking*. The method combines variation on latent semantic analysis and cocitation analysis: A document \times author matrix $\mathbf{X}(N \times P)$ and a document \times term matrix $\mathbf{Y}(N \times Q)$ are considered

$$\mathbf{ULV}^T = \text{svd}(\mathbf{Y}). \quad (5.7)$$

A vector $\mathbf{y}(N \times 1)$ is found whose rows correspond to documents that are weighted according to one or more terms. The terms are represented in a query vector $\mathbf{q}(Q \times 1)$ with $q_i = 1$ for the i ’th query terms, zero otherwise.

$$\mathbf{y} = \mathbf{U}f(\mathbf{L})\mathbf{V}^T\mathbf{q}. \quad (5.8)$$

If every principal component is maintained $f(\mathbf{L}) = \mathbf{L}$ then \mathbf{q} will just indicate the documents that contain the term. More emphasis can be added to the first principal components, e.g., with $f(\mathbf{L}) = \mathbf{L}^n$ with $n > 1$ or by setting the small singular values to zero. A principal method within a probabilistic framework would be to determine the subspace via probabilistic PCA, see section 3.9.1. After the projection further heuristic transformation can be performed on \mathbf{y} , e.g.

$$\mathbf{w} = \mathbf{y} \quad (5.9)$$

$$\mathbf{w} = \mathbf{y} > \alpha = \begin{cases} 1 & y_i > \alpha \\ 0 & y_i \leq \alpha \end{cases} \quad (5.10)$$

$$\mathbf{w} = \exp(\beta\mathbf{y}). \quad (5.11)$$

The weights in \mathbf{w} can now be applied on each document (rows) in the document \times term matrix \mathbf{X} :

$$\tilde{\mathbf{X}} = \mathbf{w}\mathbf{1}^T * \mathbf{X}, \quad (5.12)$$

where $*$ denotes the element-wise product. Finally the hub-documents and authoritative authors are found by SVD of the weighted document \times author matrix

$$\mathbf{ULV}^T = \text{svd}(\tilde{\mathbf{X}}), \quad (5.13)$$

where the most authoritative authors conditioned on the query terms are those that correspond to the highest score in \mathbf{v}_1 . A simpler and faster approach which avoids the second singular value decomposition just weight the documents with \mathbf{w}

$$\mathbf{r} = \mathbf{X}^T\mathbf{w}. \quad (5.14)$$

5.8 Example of analysis with “NeuroImage”

“NeuroImage” is a scientific journal within the functional neuroimaging field published by Academic Press (<http://www.academicpress.com/>). Electronic editions of the articles of the journal are available via the Internet

bibliographic fields could be constructed and passed between computational and visualization functions without too much bookkeeping. Not all components on the web-pages could be parsed correctly, these few examples should not affect the overall analysis.

The use of author names are not consistent: Misspellings often occur and names are represented differently, e.g. with or without middle initials. The most extreme example is to one of the most cited authors *Richard S. J. Frackowiak* whose name appears in at least 11 misspelled versions in the non-parsed corpus and 9 versions that were parsed. Examples include “Fackowiak R”, Fracknowiak R”, “Frackoviak R”, “Frackowiack R”, “Frackowiak J”, ... For this name there were 27 errors and 702 correct.

While the authors are extracted they are added to an indicator data matrix $\mathbf{X}(N \times P)$ representing N documents and P cited authors, where $x_{np} = 1$ if the n 'th document cites the p 'th author, zero otherwise. As the “cited author” the first author can be considered ($\mathbf{X}_{\text{first}}$) or all of the cited authors (\mathbf{X}_{all}). SVD can be performed on the constructed matrices

$$\mathbf{U}\mathbf{L}\mathbf{V}^T = \text{svd}_K[\mathbf{X}] \quad (5.15)$$

where the K column vectors of $\mathbf{U}(N \times K)$ and $\mathbf{V}(P \times K)$ contain weighting over N documents and P cited authors, respectively. They are here termed *eigenanthologies* and *eigencommunities*⁴. The authors in \mathbf{V} can be displayed in a plot. If the first and second principal component are used the coordinates for the p 'th author will be $(v_{p,1}, v_{p,2})$. The axes in the projection are here termed *eigenauthors*, corresponding to ideal authors that are cited in such a way the s/he only scores in one single eigencommunity, e.g., the first eigenauthor represented as a row vector $\tilde{\mathbf{u}}_{(1)} = [1, 0, 0, \dots, 0]$. Similarly, if the documents are projected the axes are termed *eigendocuments*. The Kleinberg terminology (see section 5.6.1) can be applied in connection with the first eigenauthor and first eigendocument:

- *Authorities* (authoritative sources) are the authors that score high when project onto the first eigenauthor.
- *Hubs* (hub documents) are documents that score high when projected onto the first eigendocument.

Authorities are authors that are cited by many *hub* documents and hub documents are documents that cite many authoritative authors.

For the labeling of the eigenauthor axes the titles of the citing documents are extracted, see figure 5.5. The words are split and converted to lower case with elimination of stop words. A matrix $\mathbf{Y}(N \times Q)$ is formed from the bag-of-words representation of the N documents and the Q different title words, where $y_{nq} = 1$ if the q 'th word appears in the title of the n 'th documents, zero otherwise. These words are projected onto the eigenauthor space:

$$\mathbf{Z} = \mathbf{Y}^T \mathbf{U}. \quad (5.16)$$

For the labeling of the k 'th eigenauthor axis the highest scoring words are selected from the k 'th column vector \mathbf{z}_k of the matrix $\mathbf{Z}(Q \times K)$.

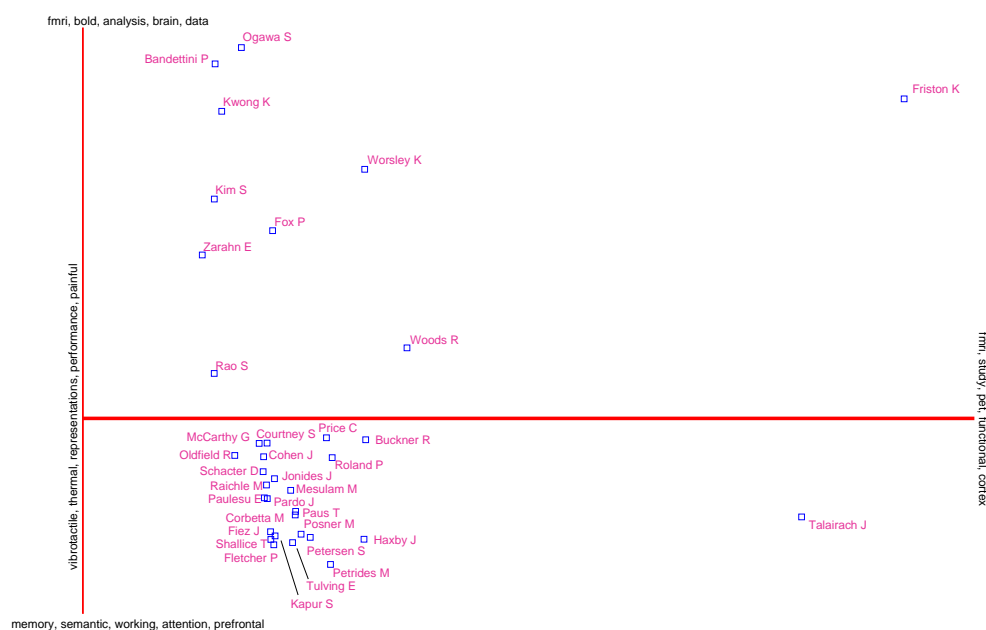
Result and discussion

Of the 325 documents in the investigated period 306 contained one or more of the 12499 references that were extracted. With the applied word stemming on the authors 4040 terms represented cited first authors and 11184 terms represented all cited authors generating (document \times cited authors) matrices $\mathbf{X}_{\text{first}}(325 \times 4040)$ and $\mathbf{X}_{\text{all}}(325 \times 11184)$. The (document \times term) matrix contained 1013 terms generated the matrix $\mathbf{Y}(325 \times 1013)$.

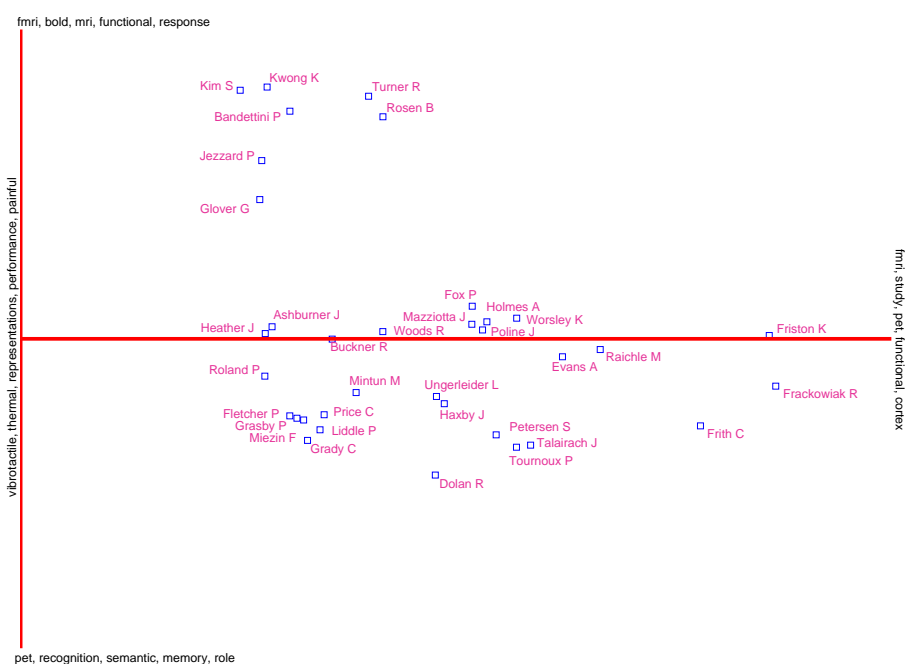
A plot of the most cited authors projected onto the two first eigenauthors with the cited first author data $\mathbf{X}_{\text{first}}$ are displayed in figure 5.6(a) with axis labeling selected as the 5 highest scoring words, — the leftmost as the highest scoring. *Karl J. Friston* and *Jean Talairach* show up as the by far most authoritative authors in the first eigenauthor. The labeling of this axis just shows the most used words in the title of the citing documents. The citation to J. Talairach is almost exclusively to the atlas books (Talairach and Szikla 1967; Talairach and Tournoux 1988). (An exception being the citation from Perani et al. 1997 to Laplane, Talairach, Meininger, Bancaud, and Bouchareine 1977). Karl J. Friston on the other hand receives citations across a broad range of his papers, though many are methodological articles associated with the SPM toolbox. The large number of citations to methodological works is also present in other fields, e.g., the paper associated with the BLAST web-service in the bioinformatics field (Altschul, Gish, Myers, and Lipman 1990) had received 12348 citations in 2001 July according to ISI® Web of Science®, and Garfield (1990) reported the article associated with the “Lowry method” (Lowry, Rosebrough, Farr, and Randall 1951) was cited the most in the period 1945–1988 with 187652 citations as recorded in the Science Citation Index. However, other methodological works become implied, and instead of receiving large number of citation they get very little cited. As an example consider spin echoes (Torben Lund, personal communication): This phenomenon was first described by Hahn (1950) and many MRI sequences rely on spin echoes. Nevertheless are there no citations to the original work in any of the NeuroImage articles.

The corresponding plot for the data with all authors are displayed in figure 5.6(b). This shows a somewhat different ordering on the authority axis. In functional neuroimaging articles often incorporate many authors. This is due to the

⁴In (Nielsen and Hansen 2001a) we used the non-intuitive terms *eigendocuments* and *eigenauthors* for the \mathbf{U} and \mathbf{V} , respectively.



(a) First cited authors



(b) All cited authors

Figure 5.6: Author cocitation analyses on cited author with citation from *NeuroImage*. The 35 most cited authors are shown in the space of the first (horizontal) and second (vertical) eigenauthor. The axes are label by the 5 words from the title of the citing documents.

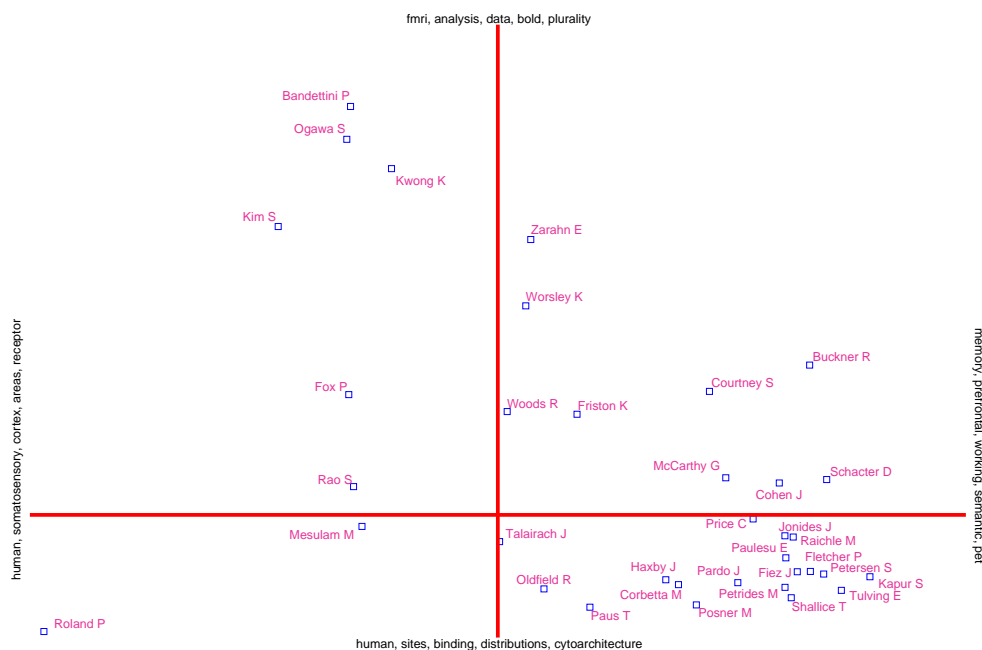


Figure 5.7: Second and third eigenauthor with idf-scaled first author matrix.

complexity and cost of a functional neuroimaging study. The last author is often a principal researcher in a grant or a head of department that authorizes the expensive scans. Richard S. J. Frackowiak is principal investigator at the Wellcome Department of Cognitive Neurological as well as previous head of MRC Cyclotron Unit at Hammersmith Hospital and has been coauthor on numerous articles from these institutions. His and Chris Frith authorities are to a large extent conditioned on the articles that Karl J. Friston has written as first author. The same applies for John C. Mazziotta and the first author Roger P. Woods. The authority for the all-author data is more related to the count from ISI® Web of Science® presented in table 5.5, e.g., Richard S. J. Frackowiak appears as the most cited author.

The second eigenauthor in both figure 5.6(a) and 5.6(b) show a segregation between fMRI and PET, and furthermore that fMRI is more associated with methodology while PET is associated with investigation of cognition — especially high-level cognitive components, e.g., memory and attention.

Other topics appear if the matrices are scaled differently and/or if higher eigenauthors are displayed. Figure 5.7 shows the second and third eigenauthor with an idf-scaled first author matrix: The third axis segregates between authors primarily associated with high and low-level cognition. A “data analysis” direction in the eigenauthor space appears as the fourth eigenauthor if the all authors are used with Poline and Worsley scoring highest.

Clustering of authors

Clustering can be performed on the cocitation data. Here non-negative matrix factorization (NMF, section 3.10) is considered: The following table appears after NMF of the document \times cited author matrix $\mathbf{X}_{\text{all}}(N \times P)$

$$\mathbf{WH} = \text{nmf}(\mathbf{X}) \quad (5.17)$$

followed by a winner-take-all on what corresponds to the eigencommunity matrix $\mathbf{H}(K \times P)$, i.e., each author is exclusively allocated to one single component.

1	2	3	4	5	6
Raichle M	Rosen B	Friston K	Haxby J	Woods R	Roland P
Petersen S	Turner R	Frackowiak R	Ungerleider L	Passingham R	Cherry S
Buckner R	Bandettini P	Frith C	Mintun M	Rao S	Mesulam M
Grasby P	Kwong K	Evans A	Petrides M	Brooks D	Toga A
Grady C	Glover G	Talairach J	D’Esposito M	Mazoyer B	Zatorre R
Miezin F	Jezzard P	Worsley K	Gore J	Binder J	Zilles K
Fletcher P	Belliveau J	Tournoux P	Noll D	Fink G	Van Essen D
Horwitz B	Kim S	Mazziotta J	Cohen J	Grafton S	Weinberger D
Posner M	Hyde J	Poline J	Goldman-Rakic P	Deiber M	Arndt S
Maisog J	Ugurbil K	Holmes A	Meyer E	Hammeke T	Gazzaniga M

The author are sorted within each of the $K = 6$ eigencommunities according to citation count and only the highest ranking are shown. As all cited authors are included the group structure will be weighted on the affiliation (through coauthorship). Some of the groups are relatively easy to interpret, e.g., the second group (the second column) contains

primarily fMRI methodology researchers and the third group comprises primarily of general methodology researchers especially researchers in connection with SPM and *Wellcome Department of Cognitive Neurology*. The other groups are harder to interpret, but the first three researchers in the first group are from St. Louis (Raichle, Petersen, Buckner), and the two first in the fourth group are affiliated with *National Institute of Mental Health* and all are known for researchers in high-level cognitive function compared to the sixth group which are more known for low-level cognitive function as well as anatomical analysis.

The interpretation of the group structure is easier when title words are projected into the groups as shown below. Here the NMF is performed on the idf-scaled first cited author matrix.

study	fmri	cortex	areas	human	fmri
pet	analysis	human	human	mri	cortex
functional	data	somatosensory	space	based	brain
fmri	plurality	receptor	stereotaxic	parcellation	functional
memory	resemblance	sites	brodmann	matter	activity
cortex	functional	regional	brought	white	analysis
activation	bold	binding	variable	cerebral	human
human	related	cytoarchitecture	brain	topographic	response
1	2	3	4	5	6
Talairach J	Worsley K	Picard N	Watson J	Pandya D	Fox P
Friston K	Zarahn E	Tanji J	Zeki S	Felleman D	Ogawa S
Woods R	Holmes A	Stephan K	Clark V	Geschwind N	Bandettini P
Roland P	Aguirre G	Penfield W	Tootell R	Van E	Kwong K
Buckner R	Boynton G	Ikeda A	Collins D	Greitz T	Kim S
Haxby J	Poline J	Deecke L	Mazziotta J	Mountcastle V	Menon R
Price C	Dale A	Bötzel K	Zilles K	Breiter H	Jezzard P
Petersen S	Cohen M	Kristeva R	Rademacher J	Kennedy D	Kleinschmidt A
Posner M	Press W	Kornhuber H	Thompson P	Alexander G	Turner R
Corbetta M	Ashburner J	Tyszka J	Duvernoy H	Caplan D	Belliveau J

The first group are primarily highly cited authors, see the authors along the first axis in figure 5.6(a). The second and the fifth group segregate fMRI, with the second group associated with analysis and the sixth group associated with acquisition, and in this group is Peter T. Fox the most cited: While he is primarily concerned with PET imaging and meta-analysis he is mostly cited for the work concerning luxury perfusion which is relevant for the understanding of the BOLD response. The third group comprises of primarily motor area researchers. The first in the list cited mostly via a review (Picard and Strick 1996). The fourth group is concerned with vision — although this is not reflected in the title words for that group. Watson, Zeki, Clark and Tootell are all cited for work within vision. Among the researchers in the fifth group are primate researchers such as Pandya, Felleman and Van Essen.

As the K columns in the eigenanthology matrix $\mathbf{W}(N \times K)$ is not necessarily orthogonal the terms can be projected with

$$\mathbf{Z}_1 = (\mathbf{W}^T \mathbf{W})^{-1} \mathbf{W}^T \mathbf{Y} \quad (5.18)$$

instead of

$$\mathbf{Z}_2 = \mathbf{W}^T \mathbf{Y}. \quad (5.19)$$

In practice with the present data set equations 5.18 and 5.19 showed little difference: the columns of \mathbf{W} are near-orthogonal.

5.8.2 Coauthor analysis

In coauthor analysis the focus is on the authors of the documents rather than the cited authors. The 325 documents have 1528 authors, and of these are 995 unique authors names with the personal name stemming used in the previous sections. There are relatively few variations in author names, — one of them being “Buechel C” and “Büchel C”, see both variations east north east in figure 5.9.

As for the author cocitation data a matrix $\mathbf{X}(N \times P)$ is constructed with $x_{np} = 1$ if the n 'th document has the p 'th author as first author or coauthor, — zero otherwise. Grouping structure in this data is much more apparent than in author cocitation data: Authors tend to write with the same group of people.

Figures 5.8 and 5.9 present bullseye visualizations of the coauthor data. Authors are connected if they share coauthorship, i.e., a line is rendered if the \mathbf{z}_{n_1, n_2} element in $\mathbf{Z} = \mathbf{X}^T \mathbf{X}$ is non-zero, n_1 representing one author and n_2 representing an other. The radial position is determined either by the number of coauthorships (figure 5.8)

$$\mathbf{r}_1 = \mathbf{X}^T \mathbf{X} \mathbf{1}, \quad (5.20)$$

or number of documents written (figure 5.9):

$$\mathbf{r}_2 = \mathbf{X}^T \mathbf{1} \underset{\text{monotone}}{\propto} \text{diag}(\mathbf{X}^T \mathbf{X}). \quad (5.21)$$

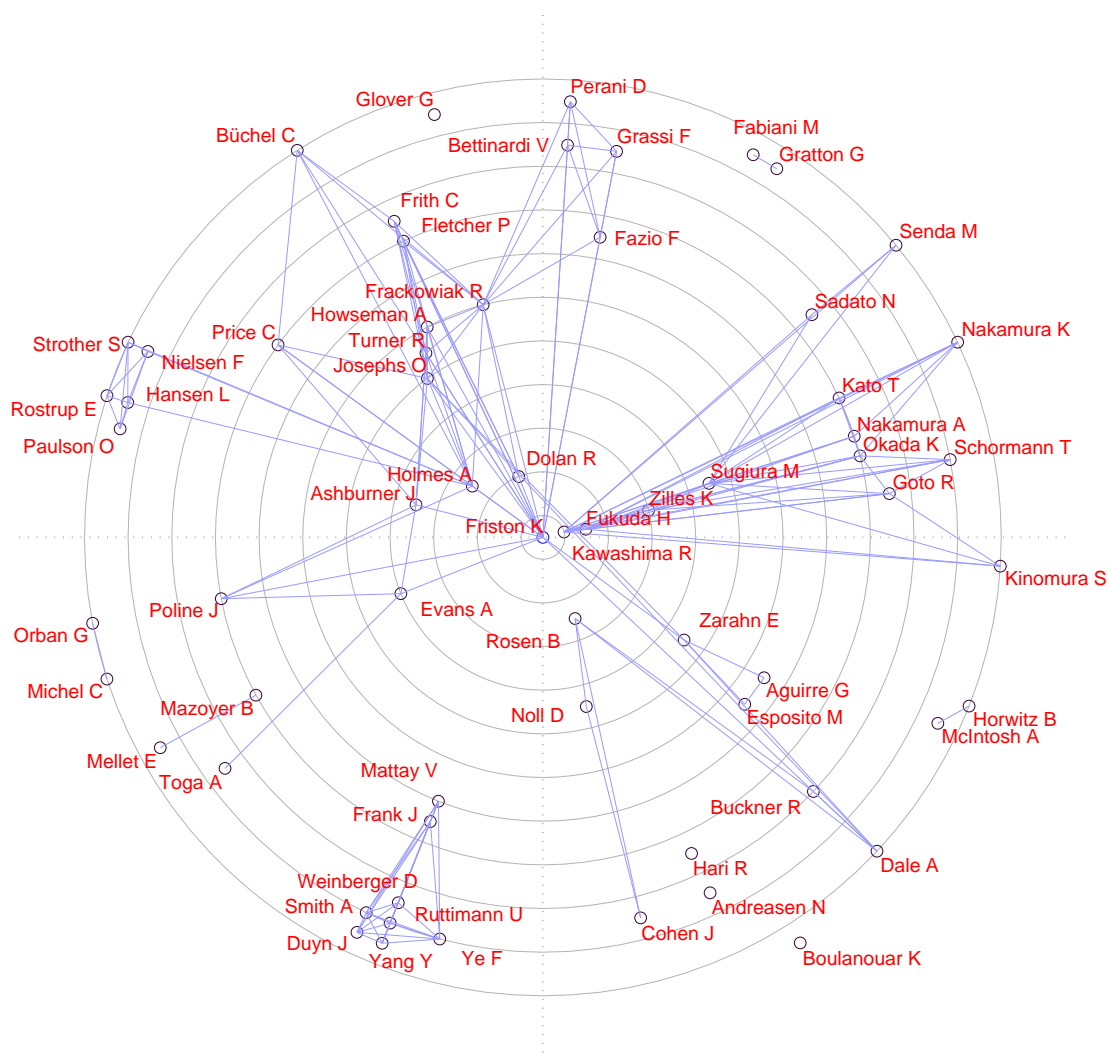


Figure 5.8: Coauthor bullseye plot displaying the 65 authors with most coauthors in *NeuroImage* in the period under study. Radial placement determined from equation 5.20. The angular placement is not optimal, e.g., Anders Dale (south east) has connection to the opposite side.

Authors with the highest rank r are near the middle and the radial metric is order rather than Euclidean. The angular placement is determined by a heuristic iterative algorithm with attraction (if connected) and repulsion. This does not optimize for the smallest number of crossing lines, but rather place nodes where there is free spaces. The visualization shows a clear group structure with the dominating group consisting of *Wellcome Department of Cognitive Neurology* led by *Karl J. Friston*. This group is connected to an Italian group and a group (our group) of primarily Danes. Separated from that is a German and Japanese group headed by *Karl Zilles* and *Ryuta Kawashima*. Other separated groups are a group from National Institutes of Health, USA, with V. S. Mattay and J. A. Frank.

A connected component analysis reveals 81 separated groups the largest groups comprising 278, 133 and 51 authors. The 10 most publishing author in the six first largest components are:

Friston K	Zilles K	Mazoyer B	Franck G	Fabiani M	Frank J
Dolan R	Kawashima R	McIntosh A	Michel C	Gratton G	Mattay V
Zarahn E	Fukuda H	Haxby J	Orban G	Alho K	Duyn J
Aguirre G	Alpert N	Mellet E	Aerts J	Corballis P	Ruttimann U
Ashburner J	Geyer S	Horwitz B	Bodart J	Kim K	Smith A
Esposito M	Sadato N	Kapur S	Bogaert P	MacIin E	Weinberger D
Holmes A	Schleicher A	Petit L	Crommelinck M	Näätänen R	Yang Y
Evans A	Schormann T	Tzourio N	Dubois S	Benveniste H	Ye F
Noll D	Sugiura M	Courtney S	Goldman S	Cowan N	Berman K
Price C	Apkarian A	Crivello F	Hecke P	Friedman D	Esposito G

When the grouping structure is so evident as the figures 5.8 and 5.9 show it is usually not a problem for unsupervised algorithm to find suitable structure in the data.

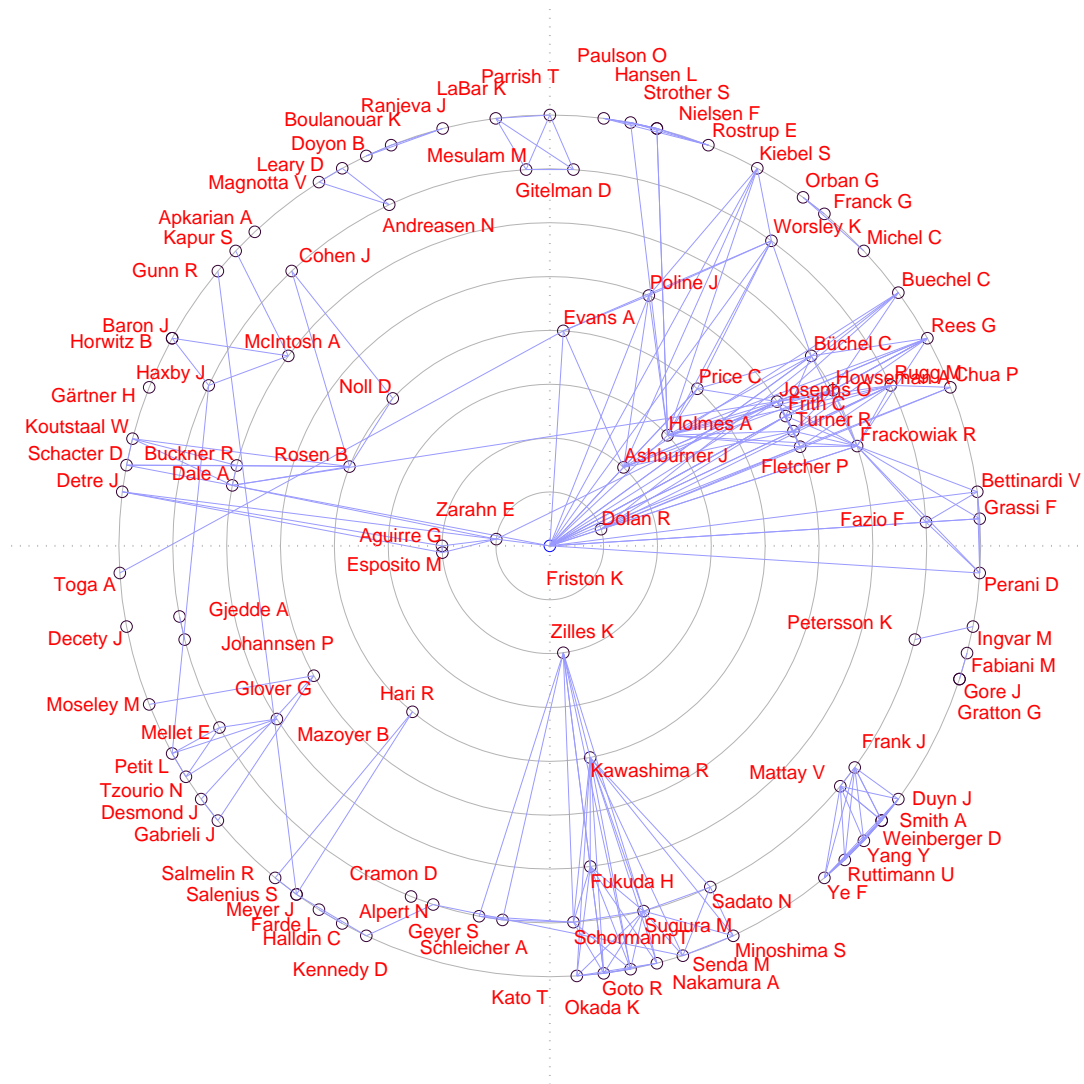


Figure 5.9: Coauthor bullseye plot with 109 of the most publishing authors in *NeuroImage* in the period under study corresponding to authors with 3 documents or more. Radial placement determined from equation 5.21.

5.8.3 Analysis of journals

As the references in *NeuroImage* also contain the title of the journal where the cited article appears the set of cited journals can be subjected to analysis. Here will *journal cocitation analysis* be considered. The analysis can help the novice in getting an overview of the field and identify key journals as well as related journals.

While the variation in author names are often little the variation in journal titles can be large. This is due to that journal titles are often abbreviated and standards (e.g., the MEDLINE abbreviation used in Pubmed) is not adhered to. As an example consider the count of references to the different variations of the “Human Brain Mapping” journal cited from *NeuroImage* listed in table 5.6(a). The most common variation has fewer counts (204) than the total of all other variations (337). Furthermore, the MEDLINE abbreviation is “Hum Brain Mapp” (without punctuation), i.e., no of the above. If the variation in journal titles are not handled both ranking and cocitation will be affected: Journals that have short non-abbreviated titles, such as “Science” or “Nature” will dominate over journals with long complex titles. Errors in journal titles are rare. The few errors are often found to be a complete word wrong, e.g., the reference to “J. Hum. Brain Map.” above and the reference to “IEEE Trans. Med. Image Anal.” — a journal that does not exist — in Royackkers et al. (1999) citation of (Vérard et al. 1997).

Based on the variations in journal naming a quick and simple word stemming method for journal titles is proposed:

1. Convert to common case (lower case). This will catch variations as “NeuroImage” and “Neuroimage”.
2. Strip punctuation “.”.

204 Hum. Brain Mapp.	1 phi tra r soc b	17 Cognit. Brain Res.
164 Hum. Brain Map.	5 phi tra r soc lon	9 Cogn. Brain Res.
101 Hum. Brain Mapping	17 phi tra r soc lon b	4 Cog. Brain Res.
33 Human Brain Mapping	3 phi tra r soc lon b bio sci	2 Brain Res. Cognit. Brain Res.
19 Human Brain Map.	2 phi tra r soc lon bio	1 Cognitive Brain Res.
7 Human Brain Mapp.	1 phi tra r soc lon ser b	1 Cognit. Brain. Res.
3 Hum Brain Map.	3 phi tra r soc lon ser b bio sci	1 Cognit. Brain Res.,
2 J. Hum. Brain Map.	2 phi tra roy soc lon ser b bio sci	1 Brain Res. Cogn. Brain Res.
2 Hum. Brain. Mapp.		
2 Hum. Brain. Map.		
1 Hum. Brain. Mapping		
1 Hum. Brain Mapping (Suppl.)		
1 Hum. Brain Mapp		
1 Hum. Brain Map		
(a) Variations of “Human Brain Mapping”	(b) Example of variation after “3-stemming”.	(c) Variation in “The Brain research. Cognitive brain research”

Table 5.6: Variation in journal naming in citations to three different journals. The journal names in the middle column have been processed with “3-stemming” described in the text.

3. Strip stop words: “on”, “the”, “of”, “and”, and sometimes: “Ser”, “Series”.

4. Word Stemming: Identify the non-abbreviated version and cut from the end of the each word. Alternatively, if the title consists of more than two words cut all words down to 3 characters, although problems arise in connection with, e.g., “phy’siology”, “phy’chiatry”, “phy’chology” and “neuro’logy” and “neuro’science”. A simple workaround is to exclude prefixes “phy” and “neuro” from the count of the word length.

Other methods would be to inquire Pubmed with the other bibliographic information (author, title, ...) and obtain the unique MEDLINE journal title string, or to enter the variation into a database.

After the 3-character stemming there can still be a large number of variations for some journals, see, e.g., Philosophical Transactions of the Royal Society, Ser B, Biological Sciences, see table 5.6(b). Another example where the presence of words in the title vary is “The Brain research. Cognitive brain research” (ISSN 0926-6410) with the MEDLINE abbreviation “Brain Res Cogn Brain Res”, see table 5.6(c). With the present word stemming the versions displayed in the table will collapse into 2 classes. If the journal impact factor is to be computed precisely then it is necessary to make more elaborate matching of journal names. Data received via ISI SCI and SSCI can also have ambiguous names (Wormell 1998, appendix).

From the 325 documents in NeuroImage 12499 references were extracted with 10795 that could be parsed as references to a journal article. There were 1209 different journal titles before and 840 after “3stemming” processing described above. These data were used to construct an indicator data matrix $\mathbf{X}(N \times P) = \mathbf{X}(325 \times 804)$ where $x_{nq} = 1$ if the n ’th document cited the q ’th journal, zero otherwise. The data matrix was subjected to singular value decomposition

$$\mathbf{ULV}^T = \text{svd}[\mathbf{X}] \quad (5.22)$$

As for author cocitation analysis column vectors of the \mathbf{U} matrix still represents a weighting over documents, so they are here termed *eigenanthologies*. The \mathbf{V} represents weighting over journals. The axes are termed *eigenjournals* when the individual journals are projected onto the principal component subspace.

Results and discussion

Simple count show that the most cited journal from *NeuroImage* is itself with over 4% of the journal citations. Furthermore, its competing functional neuroimaging sister journal *Human Brain Mapping* is the second most cited. Next in rank — as 3, 5 and 6 — are three multi-disciplinary journals with *Proceedings of the National Academy of Sciences*, *USA* (PNAS) cited more than *Nature* and *Science*. This is presumably due to that PNAS publishes more articles than the two others, — the simple count is not a measure of the *impact* since the number is not divided by the number of articles published in the journal (this is done in ordinary journal impact factor calculation), see table 5.7. The rest of the top are mostly subject-specific neuroscientific journals with *Journal of Neuroscience* the first most, and there are few review journals exceptions being *Current Opinion in Neurobiology* and *Trends in Neuroscience*. The 25 most cited journals account for almost 50% of the journal citations. Among the 25 first are none of the non-neuroscience specific biomedical journals that typically have high impact, such as *Cell*, *New England Journal of Medicine* (NEJM) and *Journal of Experimental Medicine* (Garfield 1996, figure 10), — NEJM is ranked 256 while the two others are not cited at all.

ISSN	MEDLINE abbreviation	No	ISSN	MEDLINE abbreviation	No
1065-9471	Hum Brain Mapp	50	0014-4819	Exp Brain Res	353
1053-8119	NeuroImage	124	0013-4694	Electroencephalogr Clin Neurophysiol ^a	261
0740-3194	Magn Res Med	326	0028-3932	Neuropsychologia	138
0730-725X	Magn Reson Imaging	180	0147-006X	Annu Rev Neurosci	21
0027-8424	Proc Natl Acad Sci U S A	2765	0166-2236	Trends Neurosci	125
0363-8715	J Comput Assist Tomogr	214	0036-8075	Science	1652
0271-678X	J Cereb Blood Flow Metab	148	0953-816X	Eur J Neurosci	468
0161-5505	J Nucl Med	351	0959-4965	Neuroreport	712
0364-5134	Ann Neurol	283	0959-4288	Curr Opin Neurobiol	103
0003-9942	Arch Neurol	218	0190-5295	Abstr Soc Neurosci	— ^b
0028-3878	Neurology	1054	0926-6410	Brain Res Cogn Brain Res	61
0033-8419	Radiology	568	0028-0836	Nature	1315
0278-0062	IEEE Trans Med Imaging	112	1047-3211	Cereb Cortex	88
0003-990X	Arch Gen Psychiatry	152	0898-929X	J Cogn Neurosci	51
0006-8950	Brain	353	0022-3077	J Neurophysiol	635
0006-8993	Brain Res	1370	0896-6273	Neuron	295
0304-3940	Neurosci Lett	976	0270-6474	J Neurosci	1089
0021-9967	J Comp Neurol	524			

^aNumber from year 1998. With the present word stemming it is not distinguishable from *Electromyography and clinical neurophysiology* (ISSN 0301-150X).

^bNote indexed in Pubmed

Table 5.7: Number of articles published in neuroscience journals in year 1999 obtained through queries to Pubmed with queries like “0926-6410”[Jour] AND 1999[dp]”.

The second and third principal component from an SVD with no normalization of the link matrix is shown in figure 5.10. As for author cocitation analysis of *NeuroImage* journal cocitation analysis shows a large variation can be attributed to a PET/fMRI dimension, — approximately the third principal component. The second principal axis seems to convey a microscopic/macrosopic neuroscience dimension with (macroscopic) neuroimaging in one end (Magnetic Resonance in Medicine, Journal of Cerebral Blood Flow and Metabolism, ...) and (microscopic) neurophysiology in the other (Experimental Brain Research, Journal of Neurophysiology). This is, however, not reflected in the axis labeling. Parent disciplines are partially visible in with a “medical cluster” consisting of neurology, radiology, and nuclear medicine in the fourth quadrant.

When the row sum is extracted from the link matrix a direction in the second eigenjournal appears as a statistical data analysis dimension with *NeuroImage* and *Human Brain Mapping* scoring highest in this dimension, while *Science* and *Nature* in the opposite end, see figure 5.11. This interpretation is reflected in the axis labeling with the words “data”, “statistical” and “parametric”. It reflects the low priority that *Science* and *Nature* have on methods, often confining description of methodologies to figure texts and footnotes, while *NeuroImage* and *Human Brain Mapping* often have articles that mainly focus on methodologies. A direction in fifth components has “image” and “registration” as the first two labels and *Journal of Computer Assisted Tomography* scores the highest. This is most likely driven entirely by Roger P. Woods image registration work around the popular AIR library (see sections 3.5.4-3.5.6), as the four primary references to this work is published in the journal (Woods et al. 1992; Woods et al. 1993; Woods et al. 1998a; Woods et al. 1998b).

The high loading of methodology on the second and fifth components with the row sum extracted link matrix show that the methodological articles reference fewer than neuroscientific, since the extraction of the row sum will weigh articles with few references more.

5.8.4 Finding related authors

The different cocitation matrices can be utilized in methods for determining “related articles”, “related authors” and “related journals”. Here a method for finding “related authors” is shortly described: Matrices from both first author cocitation $\mathbf{X}_{\text{first}}(N \times P)$ and all author cocitation $\mathbf{X}_{\text{all}}(N \times Q)$ are combined in a weighted addition

$$\mathbf{X} = \alpha \mathbf{X}_{\text{first}} + \beta \mathbf{X}_{\text{all}}. \quad (5.23)$$

This should be a special addition (with column permutation and zero-padding) since the P and Q columns are likely to be different. Using \mathbf{X}_{all} ensure that related authors can be found for a large number of authors and using $\mathbf{X}_{\text{first}}$ weigh the first author as most important in the citation, i.e., $\alpha > 0$. Rather than using the raw weighted sum from \mathbf{X} it is weighted by one of the forms in table 5.4: the inverse document frequency (idf), where seldom cited authors are given more weight b_q , determined as

$$b_q = \log(N/N_q), \quad (5.24)$$



Figure 5.10: Journal cocitation analysis on data from *NeuroImage*. The 35 most cited journals projected on the second and third eigenjournal.

where $N = 325$ corresponding to the total number of documents and N_q vary between 1 and $N_{\text{“Friston K”}} = 211$ for Karl J. Friston in the present data set. For the identification of “related authors” a query is formed as a vector \mathbf{x}_q picked from the q ’th column of \mathbf{X} corresponding to that q ’th author, and

$$\mathbf{r}_q = \mathbf{X}^T \mathbf{x}_q, \quad (5.25)$$

the ranking of related authors based on the q ’th author \mathbf{r}_q found as the inner product between the query vector and the weighted cocitation matrix.

Result and discussion

Results for a short list of authors are displayed in table 5.8 performed on a matrix $\mathbf{X}(N \times P) = \mathbf{X}(325 \times 11177)$ computed from equation 5.23 with $\alpha = 9$ and $\beta = 1$. For authors with very low citation counts the list is very varying depending on the precise citations in the included documents.

The result shows that with the present method the most related author is not necessarily the author himself. This would presumably change if \mathbf{X} was an abundance matrix rather than an indicator matrix. Furthermore, coauthors are not always high on the list, e.g., related authors for Karl J. Friston does not display Richard S. J. Frackowiak, which has been the last author on several of Friston’s papers.

The ranking of related authors to Ian Law is primarily determined by the citation to one of his early works (Law, Kanno, Fuhita, Miura, Lassen, and Uemura 1991) concerning reading of Japanese syllabograms: Y. Sakurai is cited by two documents as first author on documents concerning kana and kanji reading, and J. M. Rumsey has published work describing phonological and orthographic component in word reading. The others are cited only once in the corpus and are primarily researcher in neurolinguistics. Later Ian Law has published (as first author) work primarily with eye-movement activation (Law, Svarer, and Paulson 1996; Law, Svarer, Holm, and Paulson 1997a; Law, Svarer, and Paulson 1997b; Law 1996; Law, Svarer, Rostrup, and Paulson 1998) but these documents are not cited in the present corpus, and the present method fails to identify related authors in that context.

For “Nielsen F” the result shows a list of authors referenced as first authors from (Lange et al. 1999) and only from this paper (the authors have the same score and are alphabetically ordered, thus the ordering in the list is not important). Although they in this case yield somewhat interesting authors who have written documents like “Improving the precision of estimation by fitting a model”, “Publication bias: A problem in interpreting medical data”, “Independent component

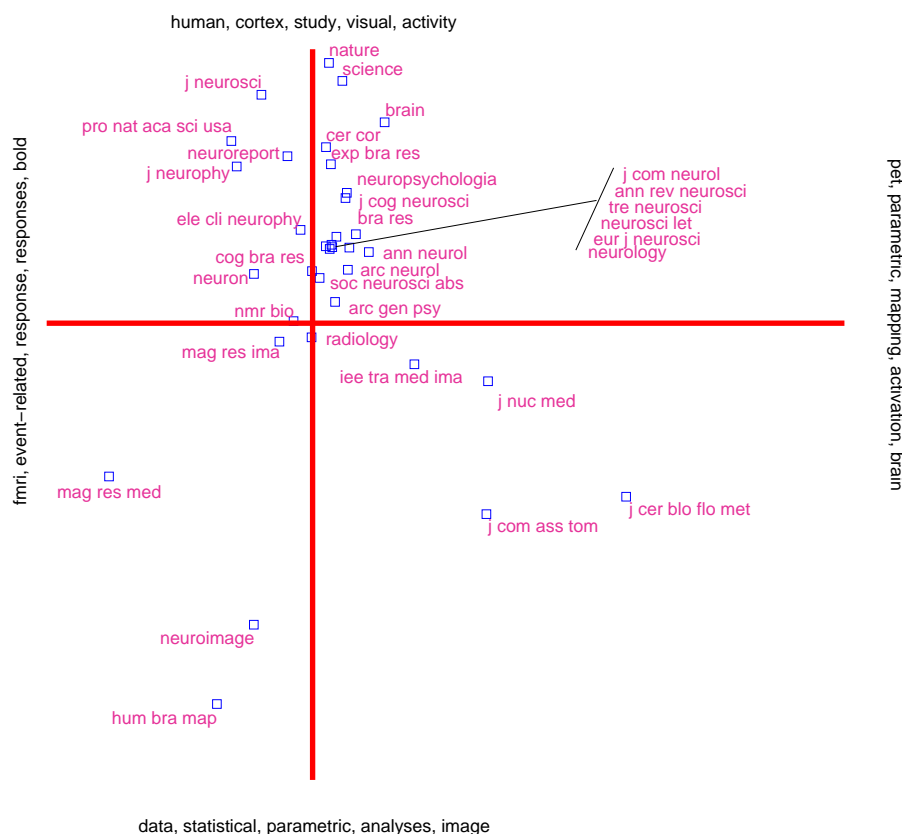


Figure 5.11: Journal cocitation analysis on data from *NeuroImage* with an the row sum.

analysis: A new concept?” which is relevant to me it might be more appropriate to disregard these single instance authors. This is done in the second column of “Nielsen F” which features results from less surprising authors mostly within pattern recognition. D. A. Leopold was not known to me and neither was Habib Benali until recently. Leopold seems to be rather unrelated while Benali has recently done work in an area related to the smooth FIR filter described in section 3.16, see (Marrelec and Benali 2001).

If the idf weighting is not applied often cited authors will move to the top of the list. For Cyril Goutte often cited researcher typically associated with fMRI acquisition move to the first and third place: Peter A. Bandettini and K. Kwong.

Extension to the scheme can include SVD on the \mathbf{X} matrix in a manner similar to latent semantic indexing, and sieving of coauthors will in most instances be preferable: An author knows his coauthors well and if he queries who is related to him he will not obtain much further information if coauthors are listed high. Also elimination of the authors own documents could be favorable or elimination of cited authors that the citing (first) author has cited.

Whether the approach described here is successful or not depend upon the goal: Should the list contain “trivial” related authors or surprising related authors. An outsider which is not familiar with an author would benefit from a list of trivial related authors, while the author himself would perhaps be more interested in viewing “novel” authors, e.g., Karl J. Friston and Lars Kai Hansen would presumably not find any novelty in their listings, — as they would know the listed authors.

Other applications apart from simple listing in connection with the authors would be email alert and conference priorities suggestion, and in this case the ranking as display in table 5.8 is interesting. E.g., I should receive an email alert when Richard Baumgartner publish a document and I should priorities his work when attending a conference (if he is attending).

5.8.5 Context sensitive author ranking

With the (document \times term) $\mathbf{Y}(N \times Q)$ and the (document \times cited author) $\mathbf{X}(N \times P)$ matrices context sensitive author ranking as described in section 5.7.1 can be performed. In the results presented below terms were projected through a 30-dimensional subspace obtained with SVD (equation 5.7).

Table 5.9 displays the ranked authors conditioned on the word “fmri”. For this term, which is well-represented in the corpus, the results are as expected: Bandettini, Ogawa, Kwong, Kim, Zarahn are all known for fMRI methodology (acquisition/analysis) most exclusively and Friston, Worsley, Buckner are known in a variety of fields and has produced key works in fMRI methodology, e.g., (Worsley and Friston 1995) and (Buckner 1998). Neither Frackowiak, Raichle

Friston K (211)	Law I (15)	Hansen L (8)	Nielsen F (4)	Nielsen F ^a (4)	Goutte C (3)	Goutte C ^b (3)
Talairach J	Sakurai Y	Hansen L	Altham P	Baumgartner R	Goutte C	Bandettini P
Friston K	Rumsey J	Akaike H	Begg C	Akaike H	Baumgartner R	Goutte C
Price C	Law I	Hastie T	Cochrane D	Duda R	Ripley B	Kwong K
Corbetta M	Coltheart V	Baumgartner R	Comon P	Hastie T	Moser E	Baumgartner R
Worsley K	Foundas A	Goutte C	Cressie N	Goutte C	Hansen L	Buckner R
Woods R	Glaser W	Ripley B	Diggle P	Ripley B	Lange N	Hansen L
Paus T	Indefrey P	Mørch N	Duhem P	Benali H	Altham P	Lange N
Martin A	Lupker S	Moser E	Friedman J	Box G	Begg C	Moser E
Ogawa S	Macleod C	Moeller J	Geary R	Cox D	Bottou L	Ripley B
Zarahn E	Makabe T	Lange N	Gnanadesikan R	Leopold D	Buckheit J	Worsley K

^aWithout authors only cited once.

^bWithout *idf* weighting.

Table 5.8: Finding related authors. The top row is the query author and the numbers in parentheses is the number of citations to that author. Below is 10 “related authors” listed with the most related on the top of the list. Parameters from equation 5.23 are set to $\alpha = 9$ and $\beta = 1$.

All authors			First authors		
No context	equation 5.13	equation 5.14	No context	equation 5.13	equation 5.14
Frackowiak R	Friston K	Friston K	Friston K	Friston K	Friston K
Friston K	Frackowiak R	Frackowiak R	Talairach J	Bandettini P	Talairach J
Frith C	Turner R	Turner R	Woods R	Ogawa S	Bandettini P
Raichle M	Rosen B	Rosen B	Buckner R	Kwong K	Worsley K
Evans A	Raichle M	Frith C	Worsley K	Worsley K	Ogawa S
Talairach J	Frith C	Holmes A	Haxby J	Talairach J	Kwong K
Worsley K	Holmes A	Raichle M	Roland P	Woods R	Buckner R
Tournoux P	Bandettini P	Bandettini P	Petrides M	Kim S	Zarahn E
Petersen S	Kwong K	Worsley K	Price C	Buckner R	Woods R
Holmes A	Worsley K	Poline J	Petersen S	Zarahn E	Boynton G

Table 5.9: Authoritative author on the term “fmri”. The second and the fifth column are constructed from equation 5.13 and the third and sixth column are constructed via equation 5.14. For comparison the first and fourth column display the ranking without weighing with terms.

nor Frith are known for work in fMRI and their appearance on the list that include all cited authors must be attributed to coauthorship.

5.8.6 Further discussion

The algorithms described in this section has been used on only a single journal, but there is little hindering that it can be used on the entire journal portfolio of a publisher — or several publishers as most of the processing is automated. Plots like the metric multidimensional plots and the bullseye visualizations can be generated once a year for each journal and display on the static HTML journal homepage along with the table of content. The enhancement needed for the processing scheme is:

- A parser that extracts the bibliographic information as well as table of content information needs to be written for each journal. In most cases a publisher would adhere to a single style so that the parser would be publisher specific. Alternatively, it should be possible to construct a general parser that can handle all common case, cf. ResearchIndex.
- The label placement is a problem with the current algorithm implemented. In the visualizations some of the point labels were moved if they were overlapping with others. An algorithm needs to be constructed that not only consider the label placement in connection with the point but also considers the placement of other points.
- The angular placement of the bullseye visualization should also be optimized.

It should be noted that documents (scientific articles) in functional neuroimaging typically have a large author list and that might be why the coauthor data shows clear group structure.

An interesting application of author cocitation analysis is ranking of emails. Emails are usually very little hyperlinked and ranking cannot be based on citation counting or the more advanced methods such as Kleinberg authorities. However,

if hyperdocuments exist that have the same authors as the emails, then the ranking from author cocitation of the hyperdocuments can be carried over to the emails. An example of this is ranking emails from the SPM-mailing list (A list associated with the SPM Matlab functional neuroimaging analysis toolbox, <http://www.jiscmail.ac.uk/lists/SPM.html>). This list contains high-quality emails from authors, which in most cases have written scientific articles that are cited. The ranking from cocitation pattern of the scientific articles can then be carried over to the SPM emails.

A note on automatic reviewer identification

Peer reviewing is an important method for distribution of grants/fellowships and for determining whether a scientific article is appropriate for publication in a journal, and inappropriate selection of reviewers can bias the result (Glantz and Bero 1994). In peer reviewing a small set of experts is selected that are to evaluate the work/application, e.g., with respect to “scientific competence”, “quality of proposed methodology” and “relevance of research project” as in (Wennerås and Wold 1997). The individual evaluations are collected and a decision is made. The grades of the reviewer are usually the only element in the review process that is quantified. An automated and quantified reviewer selection and evaluation weighting system should address the following issues:

- The reviewers should be experts in the fields.
- The reviewers should be diverse. If a work/application addresses a diverse set of fields it is not likely that reviewers can be found that each is competent in all of the fields and it would be preferable to select a set of reviewers that collectively is competent in the fields.
- The reviewers should not be affiliated with the authors/grant applicants. Wennerås and Wold (1997) were able to find correlation between affiliation and “scientific competence” as graded by the reviewers reviewing Swedish Medical Research Council postdoctoral fellowships.
- The reviewers should not be competitors of the author/applicants.
- Bias in favor of non-controversial unimportant research against innovative and controversial research, e.g., examples include the scanning tunnel microscope (Armstrong and Hubbard 1991) and MRI scanner (Kevles 1998, page 181).
- Gender bias (Wennerås and Wold 1997; Grant, Burden, and Breen 1997).
- Bias against confirmatory against negative findings.

Consider “ideal” cocitation analysis: Ranking with conditioning would be able to determine which authors are authoritative within a field, where the field is identified by its terminology. Diversity among the reviewers can be obtained by sampling among the authors so they cover a large part of the conditioned author space. Non-affiliation can be determined from the coauthor graph, while competitors would be much harder to determine. The novelty of the work/application can be assessed from the terms in the documents, though it will not be easy to determine if it describe good innovative research or just used new terms.

5.9 Example of analyses with BrainMap™

In this section analyses of the data in the BrainMap™ database (section 5.2.1) will be considered.

5.9.1 Finding outliers from the coordinates and “lobar anatomy” fields

Here probability density estimation using kernel methods (section 3.12.2) and simple term analysis (section 5.5) on BrainMap™ locations will be combined in conditional outlier/novelty detection (section 3.13). The method identifies errors in the database along with unusual terminology. The outlier detection is based on the *redundancy* in the database between anatomical labeling and the Talairach coordinates, and for this task the redundancy is crucial: “redundancy provides knowledge” (Barlow 1989) and (Hertz, Krogh, and Palmer 1991, page 197). This example is also described in (Nielsen, Hansen, and Kjems 2000; Nielsen and Hansen 2001b; Nielsen, Hansen, and Kjems 2001b).

The processing scheme is shown in figure 5.12: The “paper”, “experiment” and “locations” web-pages from the BrainMap™ Internet site are downloaded. Once the web-pages are obtained a matrix $\mathbf{Z}(N \times (3 + Q))$ is constructed from N BrainMap™ locations

$$\mathbf{Z} = [\mathbf{X}, \mathbf{Y}], \quad (5.26)$$

the matrix $\mathbf{X}(N \times 3)$ containing N Talairach coordinates from the “Coordinates in Talairach, 1988 space” BrainMap™ field and the matrix $\mathbf{Y}(N \times Q)$ containing identifiers for terms in the “Lobar anatomy” BrainMap™ text field. The Q terms are determined as the combination of all phrases (all N-grams) of the short text field with a “word” determined as a

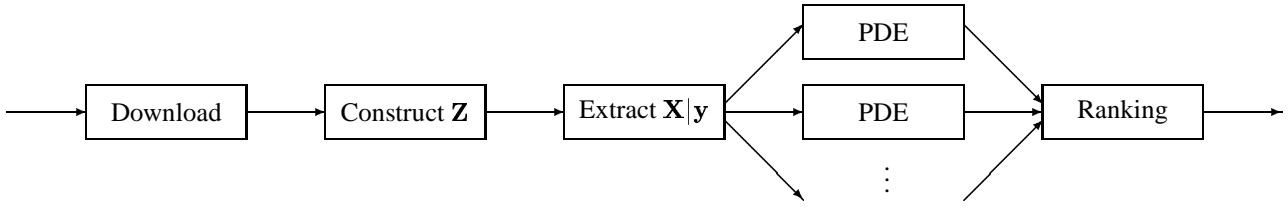


Figure 5.12: Processing scheme for finding outliers in the BrainMap™ database with downloading of web-pages, construction of matrix, extraction of submatrices, conditional probability density modeling and novelty ranking.

sequence of letters and transforming all other forms of characters to a single space (it could have been any other non-letter character):

$$[\text{^a-zA-Z}] + \longrightarrow \text{“ ”} \quad (5.27)$$

Letters are converted to lower case. The element y_{nq} is set to 1 if the n 'th location contains the q 'th term, zero otherwise.

The next step in the process is to construct $M \mathbf{X}_q$ matrices, where $M \leq Q$, e.g., if the phrase is used twice or less it is not possible to do the modeling in the proceeding steps. A novelty index is computed for each object in all the $\mathbf{X}_q (N_q \times 3)$ matrices, where $N_q = \sum_n y_{nq}$ is the number of locations containing the term q . The novelty index is determined as the probability density value of the object \mathbf{x}_{n_q} with respect to a probability density estimated for \mathbf{X}_q based on a kernel method with a homogeneous isotropic Gaussian kernel where the kernel width σ_q^2 is optimized via LOO cross-validation, equations 3.67 and 3.68. Thus the probability density estimate is conditioned on the term q

$$E(\mathbf{X}, \sigma_q^2, q) = - \sum_{n \in Q} \log p_{-n}(\mathbf{x}_n | \mathbf{X}_{Q/n}, \sigma_q^2, q) \quad (5.28)$$

$$= - \sum_{n \in Q} \log \left(\frac{1}{N_q - 1} \sum_{n' \in Q \wedge n' \neq n} (2\pi\sigma^2)^{-3/2} \exp \left[-\frac{1}{2\sigma^2} (\mathbf{x}_n - \mathbf{x}_{n'})^\top (\mathbf{x}_n - \mathbf{x}_{n'}) \right] \right) \quad (5.29)$$

where the set $Q = \{n : y_{nq} = 1\}$ is the N_q location indices containing the q 'th term and $\mathbf{X}_{Q/n}$ is a matrix of location coordinates corresponding to locations containing the q 'th term except for the n 'th location.

A two-stage strategy is used where the first step consists of modeling with all N_q objects in \mathbf{X}_q . In the second stage the 5% most extreme objects (the objects with the lowest probability density value) are discarded, and a probability density estimation is repeated with the remaining 95% objects. This corresponds to an implicit prior of 5% outliers in the sample. If there are not 5% outliers the probability density estimate will be biased. Most likely there are fewer than 5%, so the width of the probability density estimate will be smaller than the “true” width.

The process is iterated with M sets of locations each set q containing N_q locations. The entire set of location novelty $(N_1 + N_2 + \dots + N_M)$ is lastly sorted and displayed in a combined table for all M sets.

Result and discussion

The present analysis is performed on a matrix $\mathbf{X}(N \times P) = \mathbf{X}(5322 \times 1231)$, i.e., 5322 Talairach coordinates with 1231 terms (words and phrases). The terms from the anatomic labels represent a special set, that is not well associated with the normal distribution of terms in texts: The most frequent words are not the usual common words such as “the” and “of”, but “gyrus” (1st), “cortex” (2nd), “frontal” (3rd), “parietal” (4th) and “temporal” (5th). The most occurring phrases are “frontal gyrus” (12th) and “temporal gyrus” (15th) with “superior temporal gyrus” (32nd) as the most frequent 3-word phrase.

An example of a probability density volume for the “cerebellum” class is displayed in figure 5.13 in a Corner Cube environment where both the original BrainMap™ locations are displayed together with the first and second stage probability density estimate at a $P = 0.95$ threshold (0.95 of the probability mass is inside the isosurface). As expected the second-stage probability density spatially shrinks compared to the first-stage probability density. Furthermore, probability mass around outlying locations (outliers) is entirely lost in the second-stage modeling.

It is clear from the visualization that the location in the frontal cortex is a “true” outlier, and presumably also the location to the extreme right. Potentially, one could visualize all probability volumes with the associated locations and visually assess/detect the novelty of each single location. However, this would be cumbersome as there are several hundred probability volumes. The result of mathematical modeling of the locations are displayed in figure 5.14. The table displays the novelty index (log-likelihood), the BrainMap™ identifiers for paper, experiment and location (hyperlink to the BrainMap™ database) together with the Talairach coordinate and the term that the novelty is conditioned on. Also shown are the Pubmed identifiers as well as links to the full text whenever possible convenient for tracking down the cause of the novelty: The bibliographic information (authors, volume, title, journal title) on the “paper” pages can be checked against MEDLINE information via the Pubmed service on the Internet, — provided that the entry refers to a study that has been

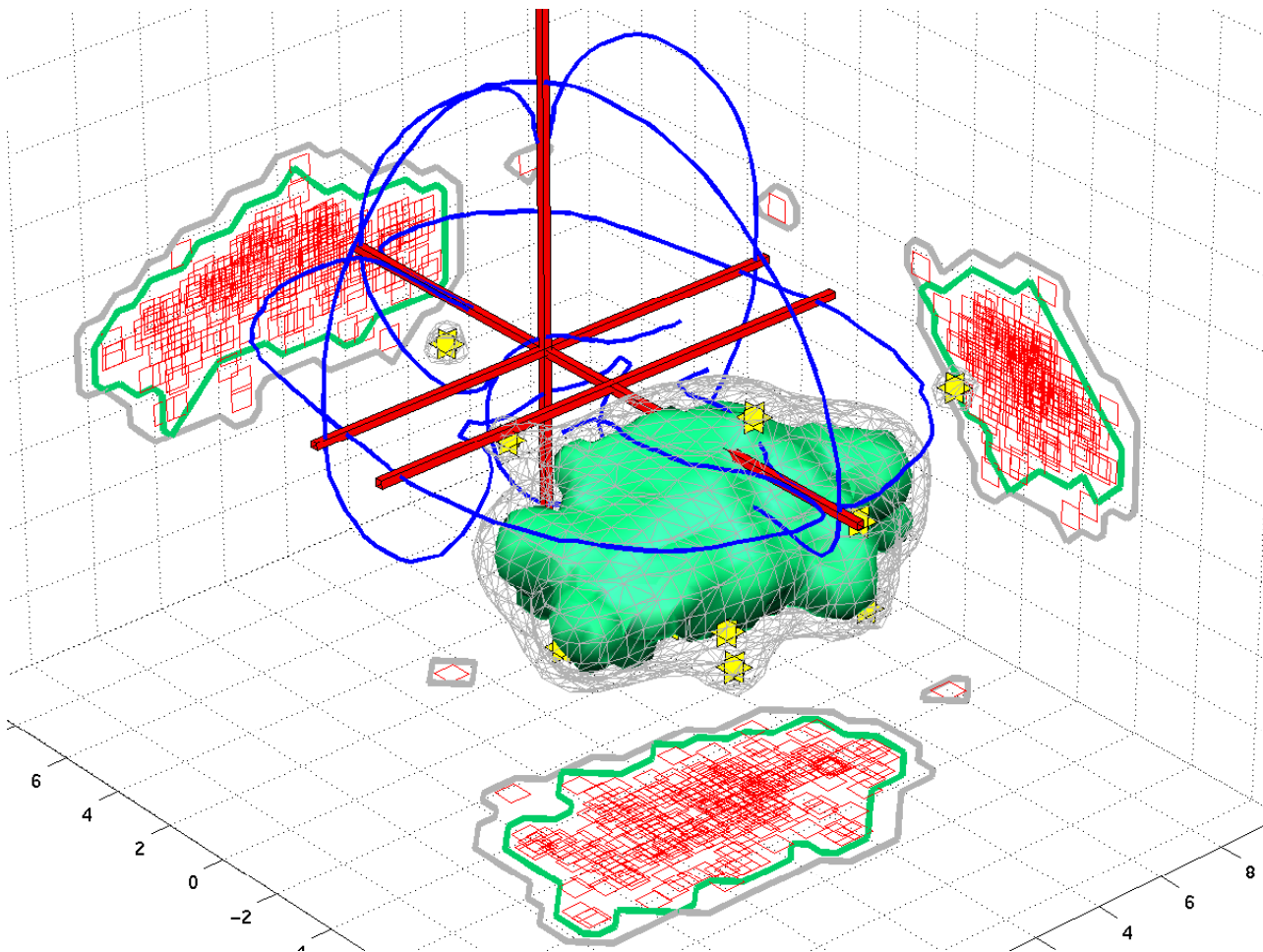


Figure 5.13: Probability density estimation of the “cerebellum” class in a Corner Cube environment. The wire-frame model is the first stage probability density estimation where all the locations are included and the polygon model is the second stage probability density estimation where the 5% most extreme are excluded. The yellow glyphs are the original BrainMap™ locations containing the term “cerebellum”. The blue curves correspond to the sagittal, coronal and transversal outline of the cortex in the Talairach atlas, and the red elongated red boxes (almost lines) are the axes in the Talairach atlas, the junctions defined at the anterior and posterior commissure. The threshold of the probability density estimate and the original BrainMap™ locations are projected orthographically onto the sagittal, coronal, and transversal walls at $(x, y, z) = (10, 10, -10)$ cm. With the displayed orientation anterior/frontal in the brain is to the far left, right in brain is to the far right and the posterior left is closest in the display.

published. A query to Pubmed that seemed to identify entries uniquely (resulted in just a single returned article) consisted of all surnames of authors, the volume of the journal, the first page and the year, — all and’ed together and marked with their type. If the query does not return a result it is either because the paper is not in the database (e.g., early papers from the journal “Human Brain Mapping” are not listed) or there is a discrepancy between the information in the BrainMap™ database and the Pubmed database. The kind of errors are, e.g., author being misspelled: “Deiner”/“Diener” (Jueptner et al. 1995), completely different: “Zeki S”/“Zatorre RJ” (Leblanc et al. 1992) or volume and number are confused. These were resolved with “manual data mining” with redundancy between databases (BrainMap™ and Pubmed) being necessary for detection.

Figure 5.14 shows the most extreme outlier as a location termed “SMA” and it clearly has a z-coordinate that is wrong: $z = 54$ cm, — half a meter outside the brain. Such an outlier would not need to be conditioned on a term to be detected as an outlier. However, an outlier that needs to be conditioned on a term in order to be detected is the seventh entry of figure 5.14. This location is inside the brain (see also figure 5.13) and is not an outlier with respect to the complete set of locations. However, when conditioning on the given label “cerebellum” it has high novelty. An example where a term provides more information than a single word are given by the second and third entries in figure 5.14 — both referring to the same BrainMap location: Adding “superior” in front of “parietal” makes the location yet more unlikely.

It can be noted that so far the analysis requires no human intervention, e.g., manual selection of the set of analyzed locations as in most current meta-analysis. To investigate the cause of the novelty it is now necessary to manually acquire,

No.	BrainMap	x	y	z	BrainMap label	Comment	Reference
1	267, 2, 1	-5	7	540	SMA	Millimeter and centimeter for z-coordinate confused during BrainMap entry	(Buckner et al. 1996, table 4, entry 1)
2	29, 10, 8	48	-23	-51	Lateral superior parietal	Resolved: Transcription mistake.	(Corbetta et al. 1993, table 5)
3	141, 1, 10	35	150	28	Dorsolateral prefrontal	Millimeter and centimeter for y-coordinate confused during BrainMap entry	(Kosslyn et al. 1994, table 2, entry 10)
4	249, 1, 59	-31.8	48.1	2.2	Subgyral frontal lobe	Correct	S. K. Brannan, 1997, Unpublished
5	280, 1, 9	24	-70	-24	Dorsal parietal cortex	Is labeled "Right cerebellum" in the article	(Schlösser et al. 1998, table 1, entry 9)
6	4, 2, 7	-6	42	-8	Cerebellum — superior anterior	Not possible to find the foci in the article.	(Petersen et al. 1988)
7	280, 1, 7	38	24	-8	Dorsolateral parietal	Is labeled "Right orbitofrontal cortex" in the article	(Schlösser et al. 1998, table 1, entry 7)
8	249,1,29	-2	26	16	Limbic Lobe	Correct	S. K. Brannan, 1997, Unpublished
9	277, 3, 3	-50	-42	-14	Inferior frontal gyrus, posterior	Is labeled "inferior temporal gyrus posterior (area 37)" in the article	(Owen et al. 1996, table 2, entry 3)
10	115, 2, 5	-38	54	0	Middle temporal gyrus	Not resolved.	(Shaywitz et al. 1995, page 155)
11	19,2,17	24	-47	38	Frontal	Not resolved	(Pardo et al. 1991, Table 1a, entry 17)
12	47,4,1	-36	32	28	Medial frontal lobe	Correct	(George et al. 1994)
13	65, 2, 23	57	26	45	Anterior cingulate	Millimeter and centimeter for x-coordinate confused during BrainMap entry.	(O'Sullivan et al. 1994, table 4, entry 10)
14	52, 1, 2	36	-46	36	Inferior frontal gyrus	Probably misunderstanding of the text during entry. The foci is around supramarginal gyrus and the denoted "BA40".	(Becker et al. 1994, page 287)
15	61, 1, 12	-24	42	4	Temporal/insular	Resolved: Transcription mistake.	(Tulving et al. 1994b, table 1)
16	130,5,8	-38	-8	4	cingulate	Perhaps a transcription error with the x- and y-coordinate being permuted	(Wills et al. 1994, table 5, entry 14)
17	48,2,3	80	-56	-16	anterior cerebellum	Millimeter and centimeter for x-coordinate confused during BrainMap entry	(Grafton et al. 1993, table 1, entry 18)
18	273,1,6	43	-14	15	parietal-occipital junction	Not resolved	(Imaizumi et al. 1997, table 1, entry 6)
19	89,1,8	-58	-37	-17	Wernicke's area	Correct, though labeled "Lt inferior temporal gyrus; Lt middle temporal gyrus (Wernicke's area)" in the article	(Leblanc et al. 1992, table 1, entry 8)
20	29,8,5	-37	-93	-8	Lingual/fusiform	Perhaps correct	(Corbetta et al. 1993, table 6)
21	26,3,4	40	-74	4	medial occipital gyrus/temporal lobe	Correct. Labeled " <i>middle</i> occipital gyrus..." in the article	(Howard et al. 1992, page 1776)

Table 5.10: BrainMap™ outliers. The entries are ordered according to novelty. The second column indicates the paper, experiment and location identifier of the BrainMap™ database. The third to fifth column are x, y and z with the "reported" coordinates from BrainMap™ (*not* the corrected "Talairach 1988" coordinates).

#	Loglikelihood	Paper	Exp	Loc	PMID	Full text	x	y	z	Lobar Anatomy
1	-Inf	267	2	1	8813903	Full text	-0.5	0.7	54.0	sma
2	-254.98	29	10	8	8441008		-4.5	-3.6	-5.4	superior parietal
3	-213.37	29	10	8	8441008		-4.5	-3.6	-5.4	parietal
4	-212.65	141	1	10	7953588		-3.5	15.0	2.8	prefrontal
5	-126.26	249	1	59			-3.2	4.8	0.2	lobe
6	-121.05	280	1	9	9576541	Full text	2.4	-7.0	-2.4	parietal
7	-120.56	4	2	7	3277066		-0.6	2.9	-0.9	cerebellum
8	-99.99	141	1	10	7953588		-3.5	15.0	2.8	dorsolateral
9	-87.58	280	1	7	9576541	Full text	3.8	2.4	-0.8	parietal
10	-81.41	249	1	29			-0.2	2.6	1.6	lobe
11	-80.71	280	1	9	9576541	Full text	2.4	-7.0	-2.4	parietal cortex
12	-78.84	277	3	3	8799180	Full text	-5.0	-4.2	-1.4	frontal
13	-66.52	115	2	5			-3.8	5.4	0.0	middle temporal
14	-61.98	19	2	17	1985266		-2.2	-6.1	4.0	frontal
15	-59.31	47	4	1			-3.6	3.2	2.8	lobe
16	-55.56	277	3	3	8799180	Full text	-5.0	-4.2	-1.4	frontal gyrus
17	-49.63	115	2	5			-3.8	5.4	0.0	temporal gyrus
18	-47.57	65	2	23	8130929		-5.7	2.6	4.5	cingulate
19	-47.12	115	2	5			-3.8	5.4	0.0	temporal
20	-46.31	52	1	2			-3.6	-4.6	3.6	inferior frontal gyrus
21	-46.04	277	3	3	8799180	Full text	-5.0	-4.2	-1.4	inferior frontal gyrus
22	-44.82	52	1	1			-4.0	-3.4	0.4	frontal
23	-42.35	52	1	2			-3.6	-4.6	3.6	frontal
24	-42.27	277	3	3	8799180	Full text	-5.0	-4.2	-1.4	inferior frontal
25	-40.68	61	1	12	8134341	Full text	-2.4	4.2	0.4	temporal

Figure 5.14: Screenshot of HTML page with the table of highest ranking novelties.

read and interpret the articles with the associated outlier locations. The result of such an analysis is shown in table 5.10 with the most novel locations from the BrainMap™ database.

Some of the novelty originate from errors and most of the errors are entry errors: Most often authors report the coordinates in millimeter, while the BrainMap™ database manipulates with the coordinates in centimeter. This difference causes a few errors during the entry into the database. In table 5.10 entries 1, 3 and 11 are examples of this. Other errors appear in the original paper: The large novelty of the second entry was due to the error in the sign of the z-coordinate: $z = -51$ should have been $z = 51$ (Personal email from Maurizio Corbetta, 2000-7-28). The 13th outlier was due to a location being mixed up with an other location. The reported coordinate $-24, 42, 4$ should be either $-42, -14, 0$ or $-44, -12, 4$ (Personal email from Endel Tulving, 2000-6-16).

The two different Talairach atlas editions (Talairach and Szikla 1967 and Talairach and Tournoux 1988) have two different orientations of the x-coordinate: Coordinates reported in the 1967 atlas system should have the sign inverted. Often authors write “left” or “right” in connection with the reported location, however, the BrainMap™ database does not maintain the words. Presumably the transformation is done automatically within the BrainMap™ entry program, leaving little risk of it being mis-entered. However, the redundancy of maintaining the label is a prerequisite for spotting left/right outliers.

Count	Lobar Anatomy
60	inferior parietal lobe
10	superior parietal lobe
6	midline occipital lobe
2	limbic lobe
1	subgyral frontal lobe
1	paracentral parietal lobe
1	medial occipital gyrus/temporal lobe
1	medial frontal lobe

Table 5.11: Count of the different number of phrases the word "lobe" appear in.

Other novelty comes from differences in terminology. E.g., the word “lobe” could denote any site in the cortex (frontal lobe, parietal lobe, ...). However, “lobe” is most often used together with “parietal”, as the counts in table 5.11 show. A probability density of such data would be focused in the parietal lobe, see figure 5.15: At the selected threshold two pronounced “blobs” appear bilaterally in the parietal lobe. A smaller blob appears in the occipital lobe — at the

midline. Several outliers are generated in connection with this term: 4th (“subgyral frontal lobe”), 8th (“limbic lobe”), 12th (“medial frontal lobe”) and 21st (“temporal lobe”).

An example on the fuzziness of labeling is the region denoted “Wernicke’s area”: It is often described as the posterior part of the left superior temporal gyrus, see, e.g., (Wise, Chollet, Hadar, Friston, Hoffner, and Frackowiak 1991). Others describe it as “left posterior superior and middle temporal gyri” (Indefrey and Levelt 2000, page 859) and (Weiller, Isensee, Rijntjes, Huber, Muller, Bier, Dutschka, Woods, Noth, and Diener 1995), “left posterior temporoparietal cortex” (Price, Veltman, Ashburner, Josephs, and Friston 1999) “temporal-occipital region” (Atkinson, Atkinson, Smith, Bem, and Hilgard 1990, page 344). Reber (1995) defines Wernicke’s areas as “a loosely circumscribed cortical area in the temporal region of the dominant hemisphere of the brain”. In the downloaded BrainMap™ database the term “Wernicke’s area” appears 5 times in the “Lobar anatomy” field, — 4 times in connection with (Kosslyn, Thompson, and Albert 1995) where the z-coordinate is either +12mm or +8mm all of the locations being labeled “superior temporal gyrus” by the Talairach Daemon, and with two locations in each hemisphere. The 19th entry table 5.10 is an outlier with respect to the term “Wernicke” and has a z-coordinate of −17mm and is labeled “inferior temporal gyrus” by the Talairach Daemon. This shows the variations in terminology on a “classic” brain area that has been known for over one hundred years.

A possible alternative scheme for detecting novelty in the BrainMap™ database would be to use anatomical atlases: Figure 5.16(a) shows the Talairach cerebellum from a triangulation of manually digitized points on the surface using the Nuages program (Geiger 1993). Many of the locations lie outside the Talairach cerebellum. Some of these should presumably not be called outliers. Other anatomical atlases are labeled probability volumes: The ICBM atlas (Evans, Collins, and Holmes 1996) being an example. In this atlas the cerebellum has been identified and each voxel is given a probability for being “cerebellum”, see figure 5.16(b). It should be noted that the volume is a *probability* volume $P(q|\mathbf{x})$, rather than a *density* volume $p(\mathbf{x}|q)$, and thus cannot be directly compared with the activation location densities. When the BrainMap™ locations are compared with respect to the probability volume it is found that some of the locations have zero probability of being “cerebellum”: $P(q = \text{“cerebellum”} | \mathbf{x}) = 0$. These are presumably not outliers. The locations in

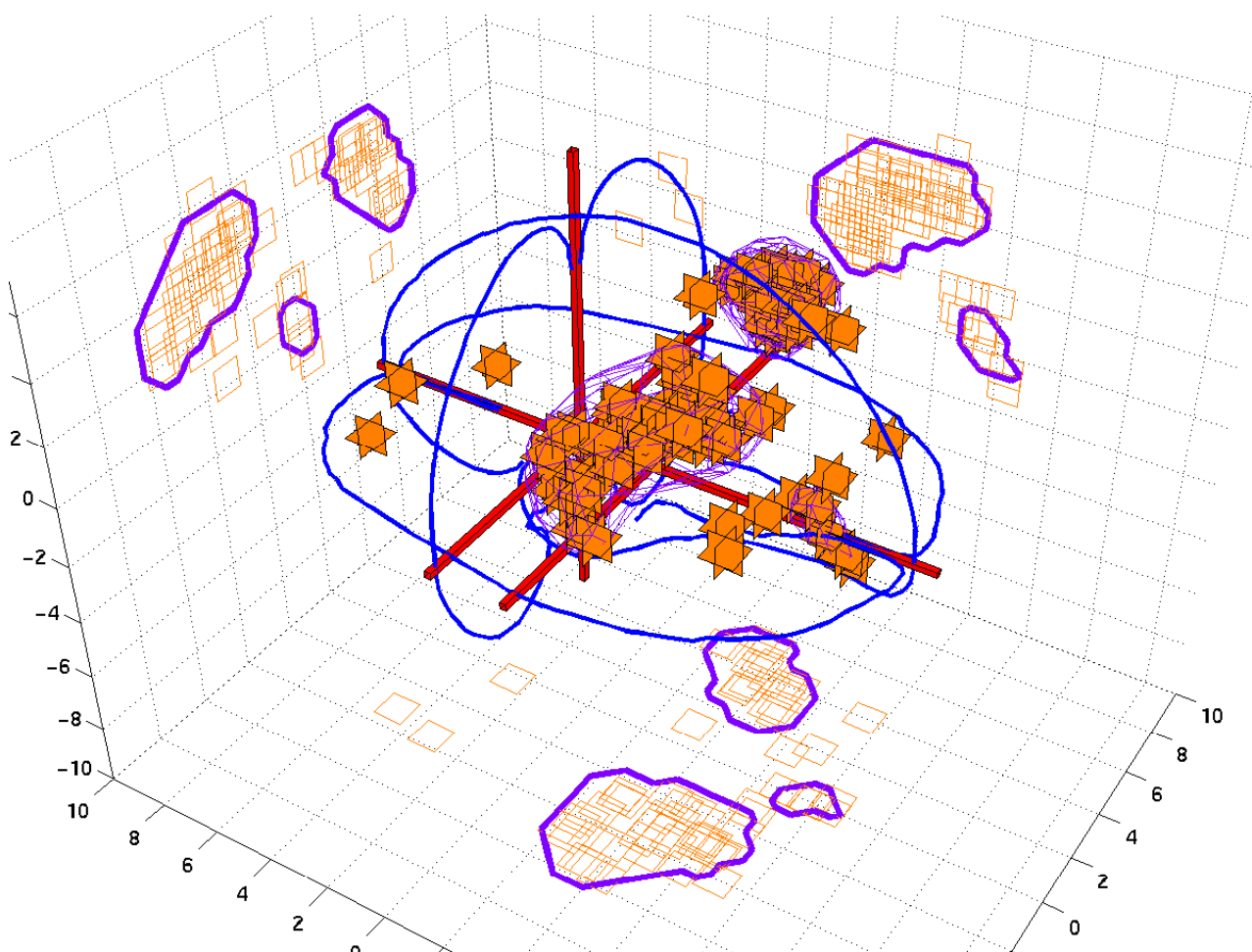


Figure 5.15: Probability density estimation of the “lobe” class shown in a Corner Cube. The blue/magenta wire-frame model is the second stage probability density estimate at a $P_{HPD} = 0.5$ threshold (highest probability density), and the orange glyphs are the “lobe” BrainMap™ locations.

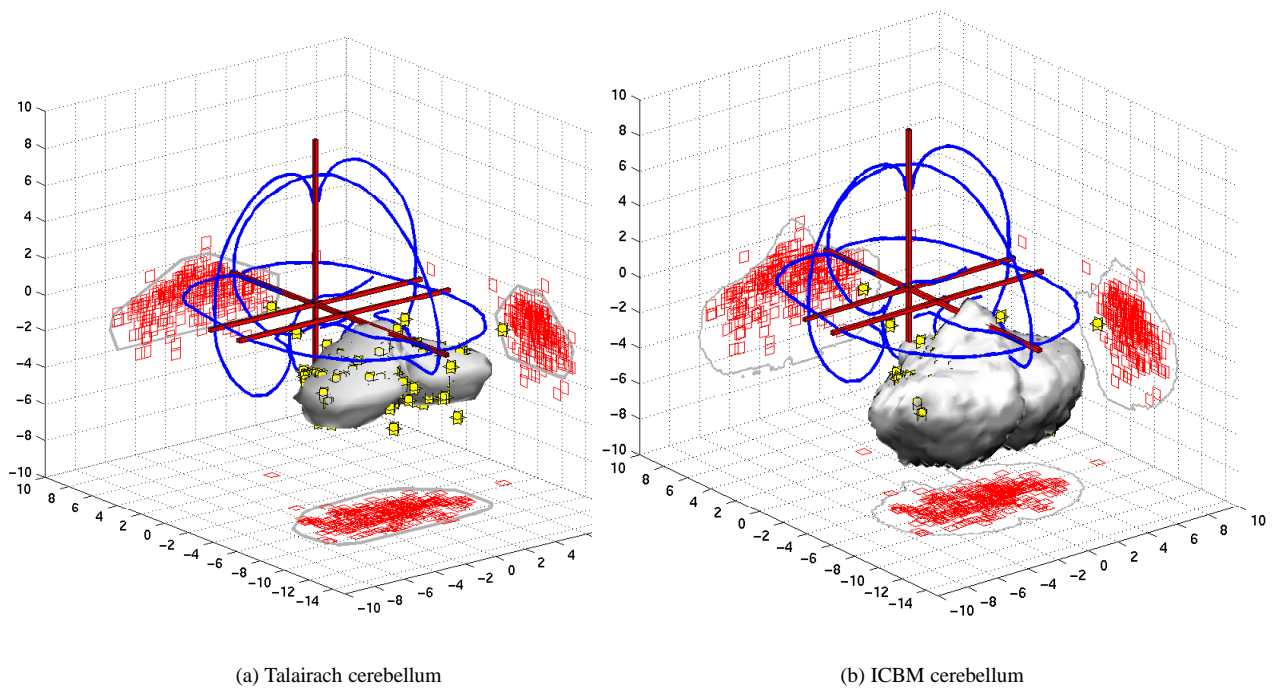


Figure 5.16: Corner Cube of “cerebellum” locations together with a digitized version of the cerebellum from the Talairach atlas and the isosurface at zero for the ICBM cerebellum probability volume. The Talairach cerebellum was digitized from the transversal slices in the atlas and polygonized by the Nuages program, see section 4.2.2. The lowest part of the cerebellum are not display in the atlas

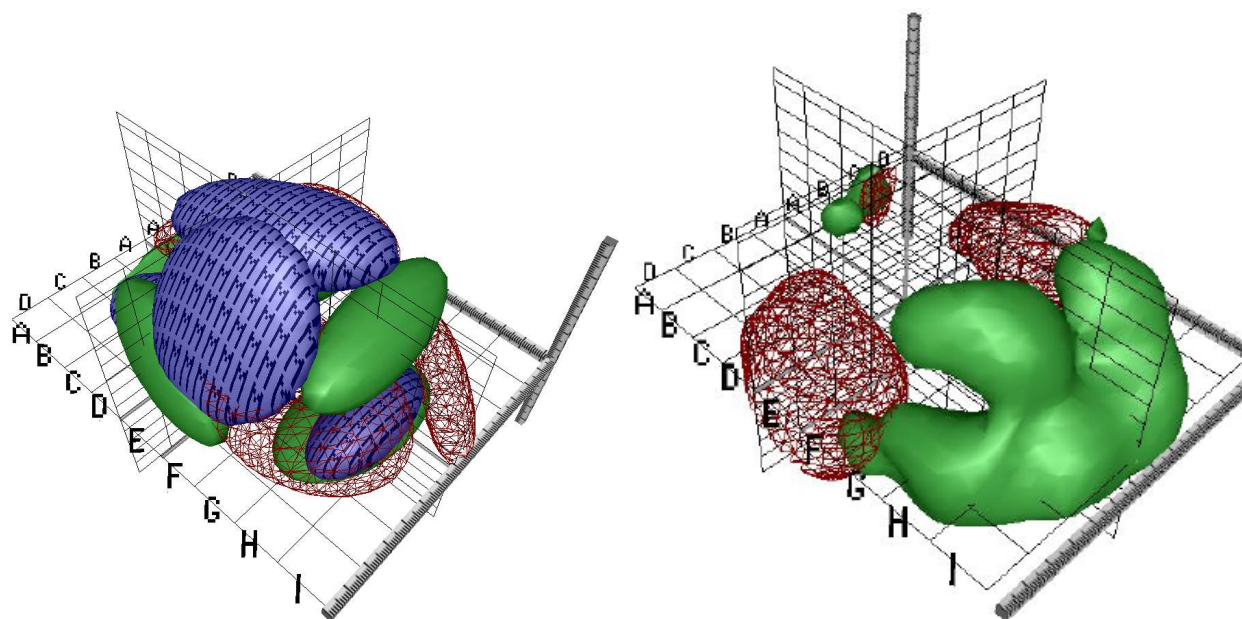
figure 5.16(b) have been transformed by the inverse operation of Matthew Brett’s nonlinear transformation (Brett 1999). It is possible that more complex spatial transformation such as a three-dimensional warps produce slightly better fit between the location and the probability volume, but probably not enough to encapsulate all of the variation in the coordinate labeling. Some variation might be attributable to the anatomical reference volume and the software used in the spatial normalization. The BrainMap™ database does not fully record this information.

5.9.2 Modeling of the functional relationship — functional volumes modeling

Cleaning of neuroscientific databases, e.g., by finding outliers in the BrainMap™ database, is important but has no direct neuroscientific value. More interesting it is to model the spatial (as e.g., represented by Talairach coordinates) and behavioral correlate, i.e., instead of conditioning on neuroanatomic labels as in section 5.9.1 the probability density estimation is conditioned on behavioral components. As already described in section 5.3.1 the method has been used by Fox et al. (1997b) with a simple parametric density estimation conditioned on mouth M1-area, and Turkeltaub, Eden, Jones, and Zeffiro (2001) that used a kernel method for modeling conditioned on a “language” behavioral component. Both these studies hand-picked the locations that were included in the study.

Figures 5.17(a) and 5.17(a) show two types of modeling with a broader set of BrainMap™ locations: Figure 5.17(a) displays 3 isosurfaces in the probability densities generated from approximately 3800 BrainMap™ locations conditioned on 3 of the 6 main behavioral domains in BrainMap™: perception, motion and cognition, cf. tables 2.1 and A.2. These densities are the posterior densities from a generalizable Gaussian mixture modeling, see section 3.12.1. Probability mass is low in the posterior part of the frontal cortex and in the lower part of the brain, e.g., inferior cerebellum. This may reflect limited field of view of the brain scanners and preference for experiments involving low-level rather than high-level cognitive components that would tend to activate the frontal cortex more. Probability is concentrated in the occipital lobe (corresponding to low-level visual activation paradigms), and in the left hemisphere — corresponding to right hand activation in primary sensorimotor areas. The components of the Gaussian mixture in this area show very little, if any, difference in the y-coordinate between motion and perception. It could have been expected that motion probability would concentrate more posterior than the perception probability, — the motor cortex being posterior to the central sulcus.

In figure 5.17(b) BrainMap™ locations are modeled with the kernel method from section 3.12.2 on locations from two categories in the second level of the behavioral domain of BrainMap™: “Vision” and “audition”, cf. tables 2.1 and A.2. Here only locations from experiments where the “Behavioral Domain” field has only one value are included and this leaves 60 audition and 366 vision locations. The isosurfaces are at $P_{\text{audition}} = 0.17$ and $P_{\text{vision}} = 0.39$. The



(a) Mixture modeling of “perception”, “cognition” and “motion”.

(b) Kernel density estimate of “Audition” and “Vision”.

Figure 5.17: Functional volumes modeling VRML screenshot. The left panel displays three of the six main categories in BrainMap™ “Behavioral Domain”: perception (red wire-frame), cognition (green polygons) and motion (blue with “M”-texture). The right panel shows a kernel density modeling of “vision” (green polygons) and “audition” (red wire-frame).

audition volume are concentrated bilaterally around the temporal lobe. The point with the highest density has Talairach coordinates $(x, y, z) = (-53, -19, 4)$ and the Talairach Daemon labels it as “Temporal Lobe” with 86.67% probability, “Left Cerebrum, Temporal Lobe, Superior Temporal Gyrus, White Matter”. A tentative hypothesis from figure 5.17(b) is that the visual activation tends to be symmetric for the ventral stream (“where” stream, location, motion action) and asymmetric left-dominated for the dorsal stream (“what” stream, form, color). Such hypotheses should be supported by a thorough literature study — and if the hypotheses are still evident experiments can be formed to test the hypotheses explicitly.

Figure 5.18 shows locations listed by Law, Svarer, Rostrup, and Paulson (1998) — a saccadic eye-movement study — colored by the probability densities displayed in figure 5.17(a), so that the red component is governed by the probability of perception, the green by cognition and blue by motion. Brightness will then be inverse proportional to an overall novelty index. The highest posterior probability for a single “behavioral domain” is the third location with the most probable label being “motion”. In the paper the location has been denoted the supplementary eye field, — an area associated with the control of eye movement. It is doubtful whether the coarse annotation (in the 3 main behavioral domain categories) helps the researcher in understanding his results. More fine grain annotation with more behavioral classes would be needed.

From figure 5.17(b) it is clear that the functional activations conditioned on such broad terms as audition and vision display a marked non-Gaussianity that can not be captured by simple standard parametric probability density models: The main characteristics of the audition volume might be captured by a relatively simple bimodal Gaussian mixture model — or a simple Gaussian model if the data is split between right and left locations, but the vision probability represents a more complex, that if modeled with a too simple model would not reveal its characteristics.

The variables that is used as conditions for the probability density volumes do not have to be in the behavioral domain, but could also be other design variables, such as modality and number of subjects.

5.9.3 Finding related experiments

How would a researcher find related functional neuroimaging experiments? The references from a “seed” article can be browsed together with its citations obtained, e.g., from ISI® Web of Science®. Another way would be to search with Pubmed using keywords and a third way would be to use BrainMap™ and search with, e.g., behavioral or location criteria. A search method that will be considered here is on the locations on the experiment level, i.e., finding related experiments determined by how similar the set of locations are.

Figure 5.19 displays a processing scheme for a simple method of finding related experiments within BrainMap™

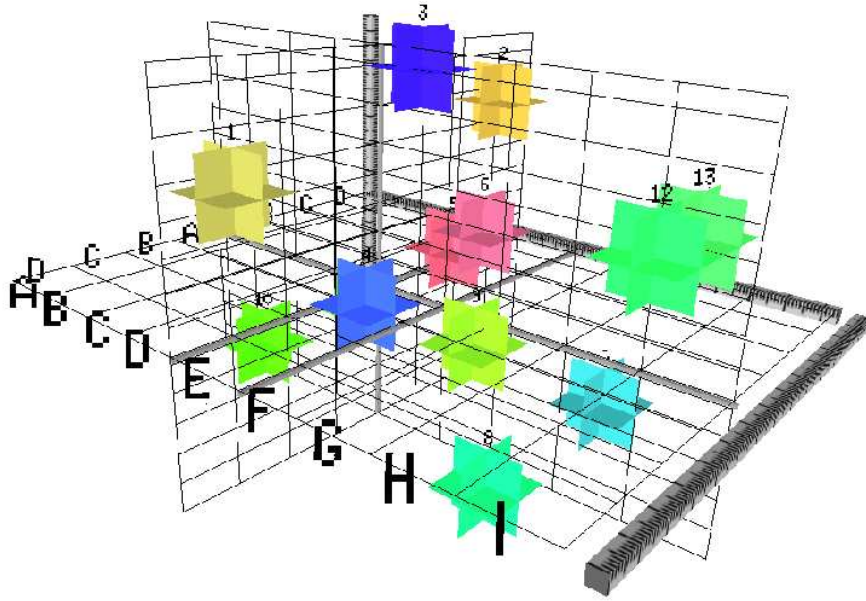


Figure 5.18: VRML screenshot of locations listed in (Law, Svarer, Rostrup, and Paulson 1998). The numbers refer to the entries in the table. The glyphs have been colored according to the posterior probability: The red component denote high probability for “perception”, the green component for “cognition” and the blue component for “motion”. Thus dark glyphs will correspond to a location with high novelty.

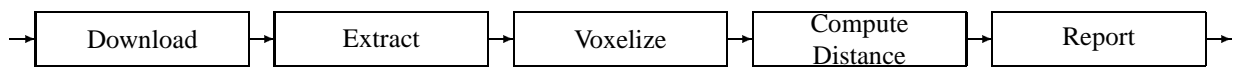


Figure 5.19: A processing scheme for finding related BrainMap™ experiments based on volume comparisons.

using volume comparison: First all BrainMap™ web-pages are downloaded and the relevant fields are extracted, i.e., most importantly the Talairach coordinates, but also other information such as bibliographic information. The sets of locations are voxelized. In the results shown below each point was convolved with an isotropic Gaussian kernel with $\sigma^2 = (0.01\text{m})^2$ and normalized with the number of location each experiment consisted of, i.e., the volume would approximately sum to one and correspond to a probability volume (since all voxels is non-negative). The volume was determined as a 8 millimeter downsampled version of the SPM99 template, see table 3.8. If the “Magnitude” BrainMap™ field of the locations is included in the voxelization negative voxels could also appear — depending on the sign of the “Magnitude”. Furthermore, the magnitudes could be used to weight the individual locations during the voxelization. When the data has been voxelized it can be represented as a matrix $\mathbf{X}(N \times P)$ with N experiments and P voxels. The distance measure is here computed as a simple inner product between a query volume \mathbf{x}_p and the volumes in the matrix \mathbf{X} .

$$\mathbf{r} = \mathbf{X}\mathbf{x}^\top. \quad (5.30)$$

This distance measure coupled with the Gaussian kernel is relatively sensitive to the spatial distance between locations. If two locations are considered and the distance between them Δx is varied then the rank/score function r will be proportional to the exponential negative squared distance $r \propto \exp[-(\Delta x)^2]$, see section B.6.1, — the value related to a cross-correlation between two signals. If instead the ranking is made inverse proportional to the squared distance $r \propto (\Delta x)^2$ then the distance is related to the differential entropy between the two voxelized volumes, see section B.6.2. In this case voxelization can be avoided and the distance measure computed as the Euclidean distance between locations.

Result and discussion

775 experiments from BrainMap™ were downloaded and 726 extracted for voxelization. With the present field of view and resolution the size of the volume was 13225 voxels, thus the resulting matrix from voxelization was $\mathbf{X}(N \times P) = \mathbf{X}(726 \times 13225)$.

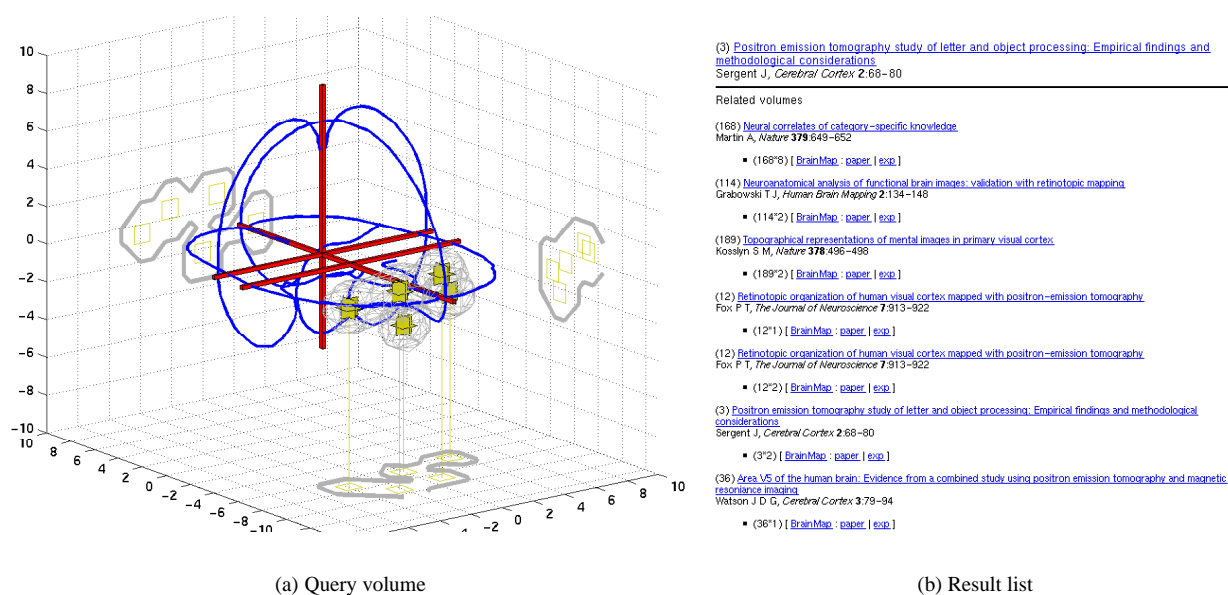


Figure 5.20: Query volume and screenshot from HTML-page with the list of results from a “find related experiments” search using BrainMap™ data.

A “query volume” is shown in figure 5.20(a) that is formed from locations from an experiment reported in (Sergent, Zuck, Levesque, and MacDonald 1992), — a letter and object (visual) processing paper. The original 5 locations are displayed as 5 glyphs and the isosurface in the volume from the voxelization is shown on the level of $P = 0.5$. Figure 5.20(b) lists the most related experiments in descending order, — most related being the eighth experiment in the paper by Martin, Wiggs, Ungerleider, and Haxby (1996) which also describes visual object processing (“naming pictures of animals and tools”). All other experiments in the displayed list are vision experiments. Another experiment from the same paper as the query experiment is ranked sixth.

There are several directions in which the approach can be extended. It would be relatively straightforward to include terms from the abstract of the BrainMap™ paper. Then a parameter determining the weight given to term versus volume should be set — either by the user or optimized in respect to some appropriate measure. Heavy weight on the terms would constitute a Pubmed-like search.

If the comparison is performed in the voxelized space not only BrainMap™ experiments can be compared but also experiments (result images) from, e.g., fMRIDC. Presently, few authors have submitted result images to fMRIDC (fMRIDC, personal communication). (Wessinger, VanMeter, Tian, Van Lare, Pekar, and Rauschecker 2001) is one of the few and this contains only single slices from individual subjects without spatial normalization, thus it is presently not possible to perform a search across databases. But as result volumes are entered in the fMRIDC database nothing should hinder the inclusion of these data. With the present size of the data static web-pages can be generated off-line and made public on a web-server. With the present resolution of the volume (13225 voxels) several thousand volumes can be searched in less than a second, thus it would be possible to make ad hoc information retrieval web-services where researchers can submit their own experiments as result volumes or location lists and get a list returned with related experiments. In this case the experiment of the researchers should be in Talairach space or a space that could be transformed to Talairach space.

If labeled data existed, e.g., from a manual scoring of related experiments, then the parameters of the search method could be optimized: kernel type, kernel width and distance measure. Other parameters that need to be set — either by optimization or user entry — are the handling of location values (the “Magnitude” of BrainMap™): Should it be ignored? Should the sign be ignored? Ordinary t -contrasts from SPM analyses yield volumes where there could both be negative and positive voxels value, — ordinary F -contrasts do only return non-negative voxels. Another option is to ignore the difference between the right and left hemisphere, i.e., ignore the sign on the x Talairach coordinate.

Chapter 6

Conclusion and discussion

This thesis has investigated some of the issues in neuroinformatics for functional neuroimaging. Much prior research in this field has focused on statistical modeling and preprocessing, and a large number of tools have been developed and still new analysis and preprocessing methods are being devised. This thesis has given an overview of some of the methods available and described a few of them in more detail — especially the unsupervised and explorative analysis methods. Furthermore, some of the multivariate analyses have been subsumed under the framework of *canonical ridge analysis*: It was shown that a type of partial least squares and orthonormal partial least squares could be regarded as regularized versions of canonical correlation analysis.

Visualization techniques suitable for multiple functional neuroimaging results in the form of both locations and volumes were proposed. A Matlab toolbox has been developed that is able to handle polygon-based graphics and output to VRML.

An overview of other areas of neuroinformatics for functional neuroimaging was presented, and since a large part of neuroscience is recorded as textual information in scientific articles an important aspect of neuroinformatics is text analysis, and this was described in detail.

Author cocitation analysis and journal cocitation analysis were shown to yield interpretable results even though the data was only obtained from a single journal and with articles from only a relatively short period. A few different visualization techniques for the bibliographic data have been suggested.

To my knowledge the analyses of NeuroImage presented in section 5.8 constitute the first citation analyses of articles from scientific journals done independently of the ISI citation indices. Obtaining citation data is in most cases restricted by the publisher. However, if publishers accept downloading of their full text items then citation analyses and visualization should be possible on a larger scale. A Matlab toolbox suitable to handle small amounts of citation data has been constructed. As most electronic versions of journals yet only cover a small number of years, this toolbox will be able to handle the bibliographic information contained in the entire corpus of a single journal in its electronic form.

Automated methods for meta-analysis of labeled locations in Talairach space were presented. The mathematical models applied were conditional probability density estimators in the form of Gaussian mixture models and kernel density estimators. Both these methods had automatic control of the generalization ability. For the kernel density estimator the kernel width was set through optimization of the cross-validation cost function and the size of the mixture model was selected by AIC while its covariance was robustly estimated by a split data set scheme.

The kernel density model was shown to be successful in connection with novelty/outlier detection in the BrainMap™ database. Probability densities in the 3-dimensional Talairach space were estimated conditioned on anatomical labels: This constitutes *conditional outlier detection*. Several outliers were found and among the top 21 the outliers constituted errors during database entry, typographical errors on the level of articles and unusual use of terminology, e.g., the term “lobe” was found to be seldomly used in connection with the “frontal lobe”, and locations labeled so would be identified as outliers.

Redundancy is a key aspect in discovering outliers. Without the redundancy between the anatomical labels and the spatial Talairach coordinate it would not have been possible to identify outliers. Ordinary database design often strives to eliminate redundancy, but here it is found that this should *not* be the case, e.g., if the words “left” and “right” are maintained in connection with the anatomical labeling of the locations, the extra redundancy provides a mean for spotting entry errors on the x Talairach coordinate. Thus a suggestion to neuroscientific database designers is not to eliminate redundancy in the entries of the database, but rather aim at describing information from a variety of perspectives and let automatic computerized mathematical models handle the relation between the information.

A method for searching in location-oriented data in Talairach space was suggested. This was based on voxelization of locations from the BrainMap™ database. It should be easy to extend to the fMRIDC database if/once this database distributes summary images in Talairach space, thus cross-database searches can be performed. Static web-pages can be constructed with “related volumes” and since it is also possible to submit entire volumes to a web-server and start a Matlab script on-line interactive search where the researcher submits s/he own volume will also be possible.

No quantitative evaluation was done in connection with the neuroinformatics methods. This was due to the high cost of generating labeled data, e.g., establishing which researchers have high authorities in, say, motor learning requires expert knowledge in that domain. The same issue arises in assessment of related volumes and the determination of the significance of the probability density volumes that are generated from functional volumes modeling. However, in connection with outlier detection in the BrainMapTM database it was possible to evaluate the performance of the method as the outliers could be checked against the information in the original article and authors could be emailed.

Several new aspects of neuroinformatics are beginning to emerge. As image databases, such as fMRIDC, make data accessible new meta-analysis and information retrieval techniques can be developed, e.g.

- In many case the data are singular, e.g., the matrix of the voxelized locations from BrainMapTM is singular as well as the matrix associated with the terms in the papers contained in BrainMapTM. *Canonical ridge analysis*, that are able to handle singular data, would be an interesting tool for characterization of the correlation between terms and anatomical location.
- In the BrainMapTM database authority is presently based on the distinction between “peer reviewed” and “unpublished”. Another type of authority is possible by using the results from cocitation with the possibility of graded authority ranked modeling and information retrieval.

Appendix A

Notation and terminology

A.1 Symbol list with main notations

Capital boldface letters are used for matrices, non-capital boldface letters for vectors, and non-boldface for scalars, e.g., matrix elements and indices.

$\|\cdot\|_F$ Frobenius norm

* Element-wise product matrix multiplication, i.e., Hadamard product, c.f. (Kiers 2000).

\otimes Kronecker product

$\mathbf{1}$ Column vector consisting of ones.

$\boldsymbol{\mu}$ (Population) mean vector

$\boldsymbol{\Sigma}$ (Population) covariance matrix.

\mathbf{B} Between groups sum of squares and products matrix

$$\mathbf{B} = \mathbf{T} - \mathbf{W} = \sum_k^K n_k (\bar{\mathbf{x}}_k - \bar{\mathbf{x}}) (\bar{\mathbf{x}}_k - \bar{\mathbf{x}})^\top \quad (\text{A.1})$$

$\mathbf{C}[\mathbf{x}, \mathbf{y}]$ The covariance matrix between \mathbf{x} and \mathbf{y} .

\mathbf{C} A Similarity matrix $\mathbf{C}(N \times N)$.

\mathbf{D} 1) A distance matrix $\mathbf{D}(N \times N)$. 2) $\mathbf{D}(P \times P)$ often denoting a diagonal matrix with standard deviations:

$$\mathbf{D} = \begin{bmatrix} s_1 & 0 & \dots & 0 \\ 0 & s_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & & s_p \end{bmatrix}. \quad (\text{A.2})$$

E Cost function

$E[\mathbf{x}]$ Expectation of a random variable \mathbf{x}

\mathbf{H} Hessian matrix $(P \times P)$ or centering matrix — usually $(N \times N)$, $\mathbf{H} = \mathbf{I} - N^{-1}\mathbf{1}\mathbf{1}^\top$, see (Mardia et al. 1979, pages 10–11)

\mathbf{I} Identity matrix, i.e., matrix with ones in the diagonal and zeros elsewhere.

K Dimension of subspace, hidden space or number of mixture components.

\mathcal{L} Likelihood

ℓ Log-likelihood

N Number of objects/examples/patterns.

P Number of variables, i.e., dimension of vector.

Q Number of variables, i.e., dimension of vector.

R Sample correlation matrix $P \times P$.

S Sample covariance matrix, $\mathbf{S} = N^{-1} \mathbf{X}^T \mathbf{H} \mathbf{X}$.

^T Matrix transpose.

T Total groups sum of squares and products (SSP) matrix

$$\mathbf{T} = \mathbf{B} + \mathbf{W} = \sum_k^K \sum_m^{N_k} (\mathbf{x}_{km} - \bar{\mathbf{x}}) (\mathbf{x}_{km} - \bar{\mathbf{x}})^T = \sum_n^N (\mathbf{x}_n - \bar{\mathbf{x}}) (\mathbf{x}_n - \bar{\mathbf{x}})^T \quad (\text{A.3})$$

U Left singular vectors (often eigensequences).

^v Matrix vectorization, i.e., stacking of the columns of a matrix into one long vector.

V Right singular vectors (often eigenimages).

$V[\mathbf{x}]$ Variance-covariance matrix ($P \times P$) of \mathbf{x} .

W Within groups sum of squares and products (SSP) matrix

$$\mathbf{W} = \mathbf{T} - \mathbf{B} = \sum_k^K \sum_m^{N_k} (\mathbf{x}_{km} - \bar{\mathbf{x}}_k) (\mathbf{x}_{km} - \bar{\mathbf{x}}_k)^T \quad (\text{A.4})$$

X Data Matrix ($N \times P$).

A.1.1 Data Matrix

The *data matrix* is convenient to represent a set of multivariate measurements. The notation from (Mardia et al. 1979, section 1.3) is used where the data matrix $\mathbf{X}(N \times P)$ is constructed by stacking N number of P -dimensional observations.

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_1^T \\ \vdots \\ \mathbf{x}_N^T \end{bmatrix} = [\mathbf{x}_{(1)} \cdots \mathbf{x}_{(P)}] = \begin{array}{c} \overbrace{\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} & \cdots & x_{1P} \\ x_{21} & x_{22} & \cdots & x_{2p} & \cdots & x_{2P} \\ \vdots & \vdots & \ddots & & & \vdots \\ x_{n1} & x_{n2} & & x_{np} & & x_{nP} \\ \vdots & \vdots & & & \ddots & \vdots \\ x_{N1} & x_{N2} & \cdots & x_{Np} & \cdots & x_{NP} \end{bmatrix}}^{P \text{ variables}} \end{array} \left. \vphantom{\begin{bmatrix} x_{11} \\ x_{21} \\ \vdots \\ x_{n1} \\ \vdots \\ x_{N1} \end{bmatrix}} \right\} N \text{ objects} \quad (\text{A.5})$$

As most multivariate analysis can be formulated with the data matrix it can be an advantage to convert a set of data into a data matrix representation: In functional neuroimaging with data from PET or fMRI scanners the data matrix can be of dimension (scans \times voxels), where the rows represent multiple measurements (scans) and the columns are voxels, which originally would have a 3-dimensional structure; in text analysis of a set of documents where a document can be converted to a *bag-of-words* that forms a vector, — with the dimension of the resulting data matrix being (documents \times words). It is not always obvious what should be the rows and columns. In functional neuroimaging a scan are usually regarded as an “observation”. However, in cluster analysis of functional neuroimaging scans it is usually the voxels that are clustered, and then it is most useful to regard a single voxel as an observation.

A.2 Word list

The word list below explain words from the areas of statistics, neural network, neuroimaging, ... For statistics the company StatSoft maintains a dictionary that is available from their homepage (<http://www.statsoft.com>). Explanation on basic statistics terms is also available from The animated software company (<http://www.animatedsoftware.com>)

For neural networks Ripley (1996) has a short glossary.

- abundance matrix:** A data matrix $\mathbf{X}(N \times P)$ that contains actual numbers of occurrences or proportions (Kendall 1971) according to Mardia et al. (1979, exercise 13.4.5).
- activation function:** The nonlinear function in the output of the unit in a neural network. Can be a threshold function, a piece-wise linear function or a sigmoidal function, e.g., hyperbolic tangent or logistic sigmoid. If the activation function is on the output of the neural network it can be regarded as a *link function*.
- active learning:** 1: The same as focusing (MacKay 1992b). 2: supervised learning Haykin (1994))
- adaptive principal components extraction:** APEX. Artificial neural network structure with feed-forward and lateral connections to compute principal components (Kung and Diamantaras 1990).
- analysis of covariance:** A type of (univariate) analysis of variance where some of the independent variables are supplementary/non-interesting — usually confounds/nuisance variable — and used to explain variation in the depended variable (Pearce 1982).
- anti-Hebbian learning:** Modeling by employing a constraint.
- asymmetric divergence:** Equivalent to the relative entropy.
- author cocitation analysis:** Analysis of the data formed when a (scientific) paper cites two different authors. The performed with, e.g., clustering technique (McCain 1990). An overview of author cocitation analysis is available in (Lunin and White 1990).
- auto-association:** Modeling with the input the same as the output.
- backpropagation:** (polysemy) The method to find the (first-order) derivative of a multilayer neural network. The method of adjusting (optimizing) the parameters in a multilayer neural network.
- bias:** (polysemy) A threshold unit in a neural network. A models inability to model the true system. The difference between the mean estimated model and the true model. The difference between the mean of an estimator and the true value.
- bias-variance trade-off:** The compromise between the simplicity of the model (which causes bias) and complexity (which causes problems for estimation and result in variance) (Geman, Bienenstock, and Doursat 1992).
- Burt matrix:** A covariance matrix of an indicator matrix (Burt 1950): $\mathbf{X}^T \mathbf{X}$
- canonical correlation analysis:** A type of multivariate analysis.
- canonical variable analysis:** A set of different multivariate analyses.
- canonical variate analysis:** Canonical correlation analysis for discrimination, i.e., with categorical variables.
- cluster:** 1: In SPM99 a region in a thresholded SPM which voxels are connected. 2: In cluster analysis a set of voxels (or other objects) characterized by a center.
- cognition:** 1: Any mental process. 2: A mental process that is not sensor-motoric or emotional. 3: The process involved in knowing, or the act of knowing (Encyclopaedia Britannica Online)
- conditional (differential) entropy:**
- $$h(\mathbf{y}|\mathbf{x}) = - \int p(\mathbf{x}, \mathbf{y}) \ln p(\mathbf{y}|\mathbf{x}) d\mathbf{x} d\mathbf{y} \quad (\text{A.6})$$
- confound:** (usual meaning) A nuisance variable that is correlated with the variable of interest.
- conjugate prior:** A prior “which have a functional form which integrates naturally with data measurements, making the inferences have an analytically convenient form” (MacKay 1995b). Used for regularization.
- consistent:** An estimator is consistent if the variance of the estimate goes to zero as more data (objects) are gathered.
- cost function:** The function that is optimized. Can be developed via maximum likelihood from a distribution assumption between the target and the model output. Other names are Lyapunov function (dynamical systems), energy function (physics), Hamiltonian (statistical mechanics), objective function (optimization theory), fitness function (evolutionary biology). (Hertz et al. 1991, page 21–22).
- cross-entropy:**
- $$- \int p(\mathbf{x}) \ln q(\mathbf{x}) d\mathbf{x} \quad (\text{A.7})$$
- Equivalent to the sum of the relative entropy and the entropy (of the distribution the relative entropy is measured with respect to). Can also be regarded as the average negative log-likelihood, e.g., with $q(\mathbf{x})$ as a modeled density and $p(\mathbf{x})$ as the true unknown density (Bishop 1995a, pages 58–59).
- cross-validation:** Empirical method to assess model performance.

dependence: In the statistical sense: If two random variables are dependent then one of the random variables convey information about the value of the other. While correlation is only related to the probability density function with the two first moments, dependency is related to all the moments.

dependent variable: The variable to be explained/predicted from the independent variable. The output of the model. Often denoted y .

design matrix: A matrix containing the “independent” variables in a multivariate regression analysis.

differential entropy: Entropy for continuous distributions

$$h(\mathbf{x}) = - \int p(\mathbf{x}) \ln p(\mathbf{x}) d\mathbf{x} \quad (\text{A.8})$$

Often just called “entropy”.

directed divergence: Equivalent to relative entropy according to (Haykin 1994).

empirical Bayes: Bayesian technique where the prior is specified from the sample information

entropy: A measure for the information content or degree of disorder.

elliptic distribution / elliptically contour distribution: A family of distribution with elliptical contours.

$$p(\mathbf{x}) \propto |\Sigma|^{-1/2} \psi[(\mathbf{x} - \mu)^T \Sigma^{-1} (\mathbf{x} - \mu)] \quad (\text{A.9})$$

The distribution contains the Gaussian distribution, the multivariate t distribution, the contaminated normal, multivariate Cauchy and multivariate logistic distribution

estimation: The procedure to find the model or the parameters of the model. In the case of parameters: a single value for each parameter in, e.g., maximum likelihood estimation, or a distribution of the parameters in Bayesian technique.

evidence: The probability of the data given the model or hyperparameters (MacKay 1992a). “Likelihood for hyperparameters” of “likelihood for models”, e.g., $p(\mathbf{X}|\mathcal{M})$.

expectation maximization: A special group of optimization methods for models with unobserved data (Dempster, Laird, and Rubin 1977).

explorative statistics: Statistics with the aim of generating hypotheses rather than testing hypothesis.

feed-forward neural network: A nonlinear model.

finite impulse response (model): A linear model with a finite number of input lags and a single output.

F-masking: Reduction of the number of voxels analyzed using a F-test.

Frobenius norm: Scalar descriptor of a matrix (Golub and Van Loan 1996, equation 2.3.1). The same as the square root of the sum of the singular values (Golub and Van Loan 1996, equation 2.5.7).

$$\|\mathbf{X}\|_F = \sqrt{\sum_n^N \sum_p^P |x_{np}|^2}. \quad (\text{A.10})$$

full width half maximum: Used in specification of filter width, related to the standard deviation by

$$\text{FWHM} = \sqrt{8 \ln 2} \sigma \approx 2.35\sigma. \quad (\text{A.11})$$

functional integration: Term used by Friston (1997a) to denote analyses of brain data with multivariate statistical methods.

functional segregation: Term used by Friston (1997a) to denote analyses of brain data with univariate statistical methods.

functional volumes modeling: A term coined by Fox et al. (1997b) to denote meta-analysis in Talairach space.

Gauss-Newton (method): Newton-like optimization method that uses the “inner product” Hessian.

general linear model: A type of multivariate regression analysis, usually where the independent variables are a design matrix.

generalized additive models: A group of nonlinear models proposed by (Hastie and Tibshirani 1990): Each input variable is put through a non-linear function. All transformed input variable is then used in a multivariate linear regression, and usually with a logistic sigmoid function on the output. In the notation of (Bishop 1995a):

$$y = g\left(\sum_i \phi_i(x_i) + w_0\right) \quad (\text{A.12})$$

generalized inverse: An “inverse” of a square or rectangular matrix. If \mathbf{A}^- denotes the generalized inverse of \mathbf{A} then it will satisfy some of the following *Penrose equations* (Moore-Penrose conditions) (Golub and Van Loan 1996, section 5.5.4) and (Ben-Israel and Greville 1980):

$$\mathbf{A}\mathbf{A}^-\mathbf{A} = \mathbf{A} \quad (I) \quad (\text{A.13})$$

$$\mathbf{A}^-\mathbf{A}\mathbf{A}^- = \mathbf{A}^- \quad (II) \quad (\text{A.14})$$

$$(\mathbf{A}\mathbf{A}^-)^* = \mathbf{A}\mathbf{A}^- \quad (III) \quad (\text{A.15})$$

$$(\mathbf{A}^-\mathbf{A})^* = \mathbf{A}^-\mathbf{A} \quad (IV), \quad (\text{A.16})$$

where $*$ denotes the conjugate, i.e., transposed matrix if \mathbf{A} is real. A matrix that satisfy all equations is uniquely determined and called the Moore-Penrose inverse and (often) denoted as \mathbf{A}^\dagger .

- generalized least squares:** 1: Regression with correlated (non-white) noise, thus were the noise covariance matrix is not diagonal, see, e.g., (Mardia et al. 1979, section 6.6.2). 2: The Gauss-Newton optimization method.
- Good-Turing frequency estimation:** A type of regularized frequency estimation (Good 1953). Often used in word frequency analysis (Gale and Sampson 1995).
- Hebbian learning:** Estimation in a model where the magnitude of a parameter is determined on how much it is “used”.
- hemodynamic response function:** The coupling between neural and vascular activity.
- heteroassociation:** Modeling where the input and output are different, e.g., ordinary regression analysis.
- hyperparameter:** A parameter in a model that is used in the estimation of the model but has no influence on the response of the estimated model if changed. An example of a hyperparameter is weight decay
- identification:** Estimation
- ill-posed:** According to Hansen (1996) a problem is ill-posed if the singular values of the associated matrix gradually decay to zero, cf. *rank deficient*.
- Imax:** Algorithms maximizing the mutual information (between outputs). Term used in Becker and Hinton (1992).
- inion:** The external occipital protuberance of the skull (Webster). Used as marker in EEG. Opposite the nasion. See also: nasion, peri-auri
- International Consortium for Brain Mapping:** (Abbreviation: ICBM) Group of research institutes. They have developed a widely used brain template know as the ICBM or MNI template.
- inversion:** In the framework of input-system-output: To find the input from the system and the output.
- Karhunen-Lóeve transformation:** The same as principal component analysis. The word is usually used in communication theory.
- kernel density estimation:** Also known as Parzen windows and probabilistic neural network.
- K-means:** Clustering technique.
- k-nearest neighbor:** A classification technique
- Kullback-Leibler distance:** Equivalent to relative entropy.
- lateral orthogonalization:** Update method (“anti-Hebbian learning”) or connections between the units in the same layer in a neural network which impose orthogonality between the different units, i.e., a kind of deflation technique. Used for connectionistic variations of singular value decomposition, principal component analysis and partial least squares among others, see, e.g., (Hertz, Krogh, and Palmer 1991, pages 209–210) and (Diamantaras and Kung 1996, section 6.4).
- learning:** In the framework of input-system-output: To find the system from (a set of) inputs and outputs. In some uses the same as estimation and training.
- leave-one-out:** A cross-validation scheme where each data point in turn is kept in the test set while the rest is used for training the model parameters.
- likelihood:** A function where the data is fixed and the parameters are allowed to vary
- $$p(\mathbf{X}|\boldsymbol{\theta}). \quad (\text{A.17})$$
- link function:** A (usually monotonic) function on the output of a linear model that is used in generalized least squares to model non-Gaussian distributions.
- lix:** Number for the readability of a text. Devised by Björnsson (1971).
- manifold:** A non-linear subspace in a high dimensional space. A hyperplane is an example on a linear manifold.
- mass-univariate statistics:** Univariate statistics when applied to several variables.
- Markov chain Monte Carlo:** Sampling technique in simulation and Bayesian statistics.
- maximal eigenvector:** The eigenvector associated with the largest eigenvector.
- maximum a posteriori:** Maximum likelihood estimation “with prior”.
- maximum likelihood:** A statistical estimation principal with optimization of the likelihood function
- m-estimation:** Robust statistics.
- Metropolis-Hasting algorithm:** Sampling technique for Markov chain Monte Carlo. (Metropolis, Rosenbluth, Rosenbluth, Teller, and Teller 1953)
- Moore-Penrose inverse:** A generalized inverse that satisfy all of the “Penrose equations” and is uniquely defined (Penrose 1955).
- multiple regression analysis:** A type of multivariate regression analysis where there is only one response variable ($\mathbf{Y} = \mathbf{y}$).
- multivariate analysis:** Statistics with more than one variable is each data set, as opposed to univariate statistics.

multivariate regression analysis: A type of linear multivariate analysis using the following model

$$\mathbf{Y} = \mathbf{XB} + \mathbf{U} \quad (\text{A.18})$$

\mathbf{Y} is an observed matrix and \mathbf{X} is a known matrix. \mathbf{B} is the parameters and \mathbf{U} is the noise matrix. The model is either called multivariate regression model (if \mathbf{X} is observed) or general linear model (if \mathbf{X} is “designed”). (Mardia, Kent, and Bibby 1979, chapter 6)

multivariate regression model: The type of multivariate regression analysis where the known matrix (\mathbf{X}) is observed.

mutual information: Originally called “information rate” (Shannon 1948).

$$I(\mathbf{y}; \mathbf{x}) = \int p(\mathbf{x}, \mathbf{y}) \ln \frac{p(\mathbf{x}|\mathbf{y})}{p(\mathbf{x})} d\mathbf{x} d\mathbf{y} \quad (\text{A.19})$$

$$= \int p(\mathbf{x}, \mathbf{y}) \ln \frac{p(\mathbf{y}|\mathbf{x})}{p(\mathbf{y})} d\mathbf{x} d\mathbf{y} \quad (\text{A.20})$$

$$= \int p(\mathbf{x}, \mathbf{y}) \ln \frac{p(\mathbf{x}, \mathbf{y})}{p(\mathbf{x})p(\mathbf{y})} d\mathbf{x} d\mathbf{y} \quad (\text{A.21})$$

nasion: The bridge of the nose. Opposite the inion. See also: inion, and periauricular points. Used as reference point in EEG.

neural network: A model inspired by biological neural networks.

neurological convention: Used in connection with transversal or axial images of the brain to denote the left side of the image is to the left side of the brain, - as opposed to the “radiological convention”.

non-informative prior: A prior with little effect on the posterior, e.g., an uniform prior or Jeffreys’ prior. Often *improper*, i.e., not normalizable.

non-parametric (model/modeling): 1: A non-parametric model is a model where the parameters do not have a direct physical meaning. 2: A model with no direct parameters. (Rasmussen and Ghahramani 2001):

[...] models which we do not necessarily know the roles played by individual parameters, and inference is not primarily targeted at the parameters themselves, but rather at the predictions made by the models

novelty: “Outlierness”. How “surprising” an object is.

nuisance: a variable of no interest in the modeling that makes the estimation of the variable of interest more difficult. See also confound.

object: A single “data point” or “example”. A single instance of an observation of one or more variables.

optimization: Used to find the point estimate of a parameter. “Optimization” is usually used when the estimation requires iterative parameter estimation, e.g., in connection with nonlinear models.

ordination: The same as multidimensional scaling.

orthogonal: For matrices: Unitary matrix with no correlation among the (either column or row) vectors.

orthonormal: For matrices: Unitary matrix with no correlation among the (either column or row) vectors.

partial least squares (regression): Multivariate analysis technique usually with multiple response variables (Wold 1975). Much used in chemometrics.

Parzen window: A type of probability density function model (Parzen 1962) where a window (a kernel) is placed at every object. The name is used in the pattern recognition literature and more commonly known as kernel density estimation.

penalized discriminant analysis: (Linear) discriminant analysis with regularization.

perceptron: A (multilayer) feed-forward neural network (Rosenblatt 1962).

periauricular: See also: nasion, inion

polysemy: The notion of a single word having several meanings. The opposite of synonymy.

preliminary principal component analysis: Principal component analysis made prior to a supervised modeling, e.g., an artificial neural network analysis or canonical variate analysis.

prediction: In the framework input-system-output: To find the output from the input and system.

principal component analysis: An unsupervised multivariate analysis that identifies an orthogonal transformation (Pearson 1901).

principal component regression: Regularized regression by principal component analysis (Massy 1965).

principal coordinate analysis: A method in multidimensional scaling. Similar to principal component analysis on a distance matrix if the distance measure is Euclidean (Mardia, Kent, and Bibby 1979, section 14.3).

principal covariate regression: Multivariate analysis technique (de Jong and Kiers 1992).

principal manifold: Generalization of principal curves, a non-linear version of principal component analysis. (DeMers and Cottrell 1993)

probabilistic neural network: A term used in Specht (1990) to denote kernel density estimation.

probabilistic principal component analysis: Principal component analysis with modeling of a isotropic noise (Tipping and Bishop 1997). The same as sensible principal component analysis.

profile likelihood: A likelihood where some of the variables — e.g., nuisance variables — are maximized (Berger, Liseo, and Wolpert 1999, equation 2)

$$\tilde{\mathcal{L}}(\theta_1) = \max_{\theta_2} \mathcal{L}(\theta_1, \theta_2). \quad (\text{A.22})$$

radiological convention: Used in connection with transversal or axial images of the brain to denote the right side of the image is to the left side of the brain, - as opposed to the “neurological convention”.

rank: The size of a subspace spanned by the vectors in a matrix

rank deficient: According to Hansen (1996) a matrix is said to be rank deficient if there is a well-defined gap between large and small singular values, cf. *ill-posed*

regularization: The method of stabilizing the model estimation.

relative entropy: A distance measure between two distributions (Kullback and Leibler 1951).

$$K(p||q) = \int p(\mathbf{x}) \ln \frac{p(\mathbf{x})}{q(\mathbf{x})} d\mathbf{x} \quad (\text{A.23})$$

Other names are cross-entropy, Kullback-Leibler distance (or information criterion), asymmetric divergence, directed divergence.

robust statistics: Statistics designed to cope with outliers.

run: A part of an experiment consisting of several measurements, e.g., several scans. The measurements in a run is typically done with a fixed frequency. Multiple runs can be part of a session and a run might consist of one or more trials or events.

saliency map: A map of which inputs are important for predicting the output.

saturation recovery: Type of MR pulse sequence.

self-organizing learning: The same as unsupervised learning

sensible principal component analysis: Principal component analysis with modeling of an isotropic noise coined by Roweis (1998). The same as probabilistic principal component analysis (Tipping and Bishop 1997)

session: A part of the experiment: An experiment might consists of multiple sessions with multiple subjects and every session contain one or more runs.

sigmoidal: S-shaped. Often used in connection about the non-linear function in an artificial neural network.

slice timing correction: In fMRI: Correction for the difference in sampling between slices in a scan. Slices can acquired interleaved/non-interleaved and ascending/descending.

softmax: A vector function that is a generalization of the logistic sigmoid activation function suitable to transform the variables in a vector from the interval $]-\infty; \infty[$ to $[0; 1]$ so they can be used as probabilities (Bridle 1990)

$$y_p = \frac{\exp(x_p)}{\sum_{p'} \exp(x_{p'})}. \quad (\text{A.24})$$

spatial independent component analysis: Type of independent component analysis in functional neuroimaging (Petersson, Nichols, Poline, and Holmes 1999a). See also temporal component analysis.

statistic: A value extracted from a data set, such as the empirical mean or the empirical variance.

statistical parametric images: A term used by Peter T. Fox et al. to denote the images that are formed by statistical analysis of functional neuroimages.

statistical parametric mapping: The process of getting statistical parametric maps: Sometimes just denoting voxel-wise t-tests, other times ANCOVA GLM modeling with random fields modeling, and sometimes also including the preprocessing: re-alignment, spatial normalization, filtering, ...

statistical parametric maps: A term used by Karl J. Friston and others to denote the images that are formed by statistical analysis of functional neuroimages, especially those formed from the program SPM.

supervised (learning/pattern recognition): Estimation of a model to estimate a “target”, e.g., classification (the target is the class label) or regression (the target is the dependent variable) estimation with labeled data.

synonymy: The notion that several words have the same meaning. The opposite of polysemy.

system: The part of the physical world under investigation. Interacts with the *environment* through *input* and *output*.

system identification: In the framework of input-system-output: To find the system from (a set of) inputs and outputs. The same as *learning*, although *system identification* usually refers to parametric learning.

temporal independent component analysis: Type of independent component in functional neuroimaging (Petersson, Nichols, Poline, and Holmes 1999a). See also spatial independent component analysis	where the diagonal \mathbf{L}_K is $[l_1, l_2, l_3, 0, \dots, 0]$. The truncated SVD matrix is the K -ranked matrix with minimum 2-norm and Frobenius norm of the difference between all K -ranked matrices and \mathbf{X} .
test set: Part of a data set used to test the performance (fit) of a model. If the estimate should be unbiased the test set should be independent of the training and validation set.	univariate statistics: Statistics with only one response variable, as opposed to multivariate analysis.
time-activity curve: The curve generated in connection with dynamic PET images	unsupervised learning: Learning with only one set of data, — there is no target involved. Clustering and principal component analysis is usually regarded as unsupervised.
total least square: Multivariate analysis estimation technique	voxel: A 3-dimensional pixel. The smallest picture element in a volumetric image.
training: Term used in connection with neural networks to denote parameter estimation (parameter optimization). Sometimes called learning.	validation set: Part of a data set used to tune hyperparameters.
training set: Part of a data set used to fit the parameters of the model (not the hyperparameters). See also test and validation set.	weights: The model parameters of a neural network.
trial: An element of a psychological experiment usually consisting of a stimulus and a response.	white noise: Noise that is independent (in the time dimension).
truncated singular valued decomposition: Singular value decomposition of a matrix $\mathbf{X}(N \times P)$ where only a number of components, say $K < \text{rank}(\mathbf{X})$, are maintained	z-score: Also called “standard score” and denotes a random variable transformed so the mean is zero and the standard deviation is one. For a normal distributed random variable the transformation is:

$$\hat{\mathbf{X}}_K = \mathbf{U}\mathbf{L}_K\mathbf{V}^T, \quad (\text{A.25})$$

$$z_x = \frac{x - \mu_x}{\sigma_x} \quad (\text{A.26})$$

A.3 General Abbreviations

This list contains acronyms within the areas of statistics and functional neuroimaging. All the abbreviation from (Law 1996) is listed. The Acrophile web-service provides more general types of acronyms (Larkey et al. 2000).

ACA: Author cocitation analysis	AR: auto regressive (model).
ACF: Autocorrelation function	ARMA: autoregressive moving average (model).
AIC: Akaike’s information criterion (or “An information criterion)	ARMS: Adaptive rejections Metropolis sampling.
AIR: Automated image registration (program and method for aligning neuroimages)	ARS: Adaptive rejection sampling. Sampling technique.
ANCOVA: Analysis of covariance.	ASCII: American Standard Code
ANOVA: Analysis of variance.	BA: 1: Brodmann area. 2: Broca’s area.
AR: Autoregressive	BGO: Bismuth germanate oxide
ARMA: Autoregressive moving average.	BIC: Bayesian information criterion
ARMAX: Autoregressive moving average, exogenous (inputs).	BIC: Information criterion B
aMRI: Anatomical magnetic resonance imaging (or image). Used to distinguish it from an fMRI.	BOLD: Blood oxygen level-dependent.
ANOVA: Analysis of variance.	Bq: Becquerel
APEX: Adaptive principal components extraction.	CBF: Cerebral blood flow.
	CCA: Canonical correlation analysis
	CGI: Common gateway interface
	Ci: Curie

COG:	Center of gravity	HME:	Hidden Markov Model.
CSF:	Cerebral spinal fluid.	HPD:	Highest posterior density.
CT:	Computer tomography	HRF:	Hemodynamic response function.
CVA:	Canonical variate analysis.	HTML:	Hypertext markup language.
DAG:	Directed acyclic graph.	HTTP:	Hypertext transfer protocol.
DBM:	Deformation-based morphometry.	ICA:	Independent component analysis
DOF:	Degrees of freedom.	ICBM:	International Consortium for Brain Mapping
DTI:	Diffusion tensor image.	idf:	Inverse document frequency
DWI:	Diffusion weighted image.	IIR:	Infinite impulse response.
ED:	Effective Dose	iOIS:	Intraoperative optical imaging of intrinsic signals.
EEG:	Electroencephanography or electroen- cephanogram	IR:	1: Information retrieval. 2: Intraoperative in- frared (imaging). 3: Inversion recovery. Type of MRI pulse sequence.
efMRI:	Event-related functional Magnetic Resonance	IRF:	Impulse response function
ENG:	Electronystagmography	ISI:	1: Institute for Scientific Information. 2: Inter- stimulus intervals. 3: Interscan Interval.
EMG:	Electromyogram.	ITI:	Inter-trial interval
EOG:	Electro-oculography.	LED:	Light emitting diode
EPI:	Echo-planar imaging. Type of MRI pulse se- quence. Used in fMRI.	LMS:	Least mean square
erfMRI:	Event-related functional magnetic resonance imaging. Also efMRI.	LON:	Lateral orthogonalization network
EM:	Expectation Maximization	LOO:	Leave-one-out
ESM:	Electrocortical Stimulation mapping.	MAP:	Maximum a posteriori
FA:	Flip angle.	MANCOVA:	Multivariate analysis of covariance
FAIR:	Flow-Sensitive Alternating Inversion Recovery	MANOVA:	Multivariate analysis of variance.
FDG:	Fluorodeoxyglucose ^{18}F -2fluoro-2-deoxyglucose	MCMC:	Markov chain Monte Carlo.
FID:	Free induction decay.	MDL:	Minimum description length.
FIR:	Finite impulse response.	MDS:	Multidimensional scaling
fMRI:	Functional magnetic resonance imaging (or im- age).	MEG:	Magnetoencephalography.
FOV:	Field of view. The dimension of the scanner im- age.	MLE:	Maximum likelihood estimator.
FVM:	Functional volumes modeling.	MRF:	Markov random field.
FWHM:	Full width half max.	NCS:	Nerve conduction
GE:	Gradient echo	NMR:	Nuclear magnetic resonance.
GLM:	1: General linear model. 2: Generalized linear models.	OIS:	Optical imaging of intrinsic signals.
GLMM:	Generalized linear mixed models. Generalized linear models with random effects.	PACF:	Partial auto-correlation function.
GLS:	Generalized least squares	PCA:	Principal component analysis.
H2-15O:	Oxygen-15 labeled water.	PCoA:	Principal coordinate analysis
		PCR:	Principal component regression.
		pdf:	probability density function

PDF:	Portable Document Format	SNR:	Signal-to-noise ratio.
PET:	Positron Emission Tomography.	SOM:	Self-organizing map.
PLS:	Partial least squares (also called projected latent structures).	SOA:	Stimulus onset asynchrony
PLSA:	Probabilistic latent semantic analysis	SPCA:	Sensible Principal component analysis
pMRI:	Perfusion based MRI.	SPI:	Statistical parametric images
PPCA:	Preliminary principal component analysis	SPM:	1: Statistical parametric mapping. 2: Statistical parametric maps. 3: A program for analysis of functional neuroimages.
PMT:	Photomultiplier tube	SR:	Saturation Recovery. Type of MRI pulse sequence.
PSTH:	Poststimulus Time Histograms (neurophysiology)	Sv:	Sievert
rCBF:	Regional cerebral blood flow	SVD:	Singular value decomposition.
relCBF:	Relative cerebral blood flow	SVC:	Small volume correction
RF:	Radio frequency. The pulse that is used in MRI to excite the nucleus of the atom.	TAC:	Time-activity curve
ROI:	Region of interest.	TE:	Echo time
MRI:	Magnetic resonance imaging (or image)	tICA:	temporal independent component analysis
MRS:	Magnetic resonance spectroscopy	TLS:	Total least square
rCBF:	Regional cerebral blood flow	TR:	Repetition time. Repeat time
rCBV:	Regional cerebral blood volume	TSVD:	Truncated singular value decomposition
rCMR:	Regional cerebral metabolism rate	URL:	Uniform resource locator
rCMRglu:	Regional cerebral metabolic rate of glucose	VAC:	Vertical projection of anterior commissure
rCMRO2:	Regional cerebral metabolic rate of oxygen	VBM:	Voxel based morphometry
rcounts:	Regional counts	VOI:	Volume of Interest
rOEF:	Regional oxygen extraction fraction	VPC:	Vertical projection of posterior commissure
SE:	Spin Echo.	VRML:	Virtual Reality Modeling Language
sICA:	spatial independent component analysis		
S/N:	Signal-to-Noise		

A.4 Anatomical names and abbreviations

anterior:	Frontal.	lateral:	Toward the side. Opposite medial.
caudal:	Referring to a direction toward the tail (of the spinal cord). Opposite rostral.	medial:	Toward the middle. Opposite lateral.
contralateral:	Opposite side. Opposite ipsilateral.	posterior:	Back. Sometimes used for caudal or dorsal.
dorsal:	Relating toward the back. Opposite ventral.	rostral:	Referring to the part of the spinal cord (and brain) toward in the nose. Opposite caudal. Somehow equivalent: superior, anterior, ventral.
fundus:	Bottom.	superior:	Top or above.
inferior:	Bottom or below. Opposite superior.	ventral:	Referring to the “belly-side” (frontal part) of the spinal cord. Opposite dorsal.
ipsilateral:	Same side. Opposite contralateral.		

No.	Sub	Location	Function
1		(Intermediate) postcentral gyrus	Somatosensory
2		(Caudal) postcentral gyrus	Somatosensory
3		(Rostral) Postcentral gyrus	Somatosensory
4		Frontal lobe, precentral gyrus, Gigantopyramidal	(Primary) motor cortex
	4a	Anterior	Self generated movements(?)
	4p	Posterior	Roughness discrimination(?)
5		Superior parietal lobule, preparietal	
6		Frontal lobe / Precentral gyrus, agranular frontal	Part of Premotor cortex
7		Superior parietal lobule / precuneus	
8		Mesial frontal cortex, intermediate frontal	Frontal eye field
9		Superior frontal gyri / (medial) prefrontal cortex, granular frontal	
10		Anterior prefrontal cortex / medial frontal gyrus, frontopolar	Cognition(?)
11		Orbitofrontal cortex, prefrontal	Cognition(?)
12		Prefrontal	Cognition(?)
13		(Only in monkey)	
14		(Only in monkey)	
15		(Only in monkey)	
16		(Only in monkey)	
17		Occipital lobe, striate	Vision, primary visual cortex
18		Occipital lobe, extra striate, parastriate	Vision, secondary visual cortex
19		Occipital lobe, peristriate	Vision
20		Inferior temporal	Visual
21		Middle temporal	Visual
22		Temporal lobe, posterior superior temporal gyrus, superior temporal	(Spoken) language, Wernicke
23		Posterior cingulate, ventral posterior cingulate	Cognition(?)
24		Anterior cingulate gyrus, ventral anterior cingulate	Cognition(?), autism ^a , attention ^b
25		Inferior antero-caudal cingulate gyrus, subgenual	Cognition(?)
26		Ectosplenial	Cognition(?)
27		(Only in monkey)	
28		Amygdala, entorhinal	
29		Granular retrolimbic	
30		Retro-splenial region, agranular retrolimbic	
31		Posterior cingulate gyrus	
32		Mesial frontal cortex / Anterior cingulate	Motor
33		Pregenua	Cognition(?)
34		Parahippocampal gyrus, dorsal entorhinal	Olfaction
35		Perirhinal cortex	Cognition(?)
36		Ectorhinal cortex	Cognition(?)
37		Posterior fusiform, occipitotemporal	Visual
38		Superior temporal / Inferior temporal pole, temporopolar	Emotion
39		Angular gyrus, parieto-occipito-temporal junction	(Perception of written) language
40		Supramarginal gyrus / parietal lobe	Somatosensory(?)
41		Temporal lobe, Heschl's gyrus, anterior transverse temporal	Hearing
42		Temporal lobe, Heschl's gyrus, posterior transverse temporal	Hearing
43		Subcentral	
44		Inferior frontal gyrus, opercular	Speech, Broca
45		Inferior frontal gyrus, triangular	Speech, Broca
46		Prefrontal cortex, middle frontal	Cognition(?), response selection ^c
47		Insula / Inferior frontal gyrus, orbital	Cognition(?)
48		Retrosubicular	
49		(Only in monkey)	
50		(Only in monkey)	
51		(Only in monkey)	
52		Parainsula	

^aHaznedar, Buchsbaum, Wei, Hof, Cartwright, Bienstock, and Hollander (2000)^bBush, Frazier, Rauch, Seidman, Whalen, Jenike, Rosen, and Biederman (1999)^c(Rowe et al. 2000)

Table A.1: Brodmann numbers. Some of the information in this table was taken from web-pages of Mark Dubin (University of Colorado) <http://spot.colorado.edu/~dubin/talks/brodmann/>, NeuroNames (Bowden and Martin 1995) and (Heimer 1994). The function column is tentative.

Type	Subtype	Example
Perception	Audition	Lexical coding Noise Speech Words
	Gustation Olfaction Somethesis	Pain Temperature Vibration
Motion	Vision	Color Letters Letter recognition Motion Objects Patterened stimulation Photic stimulation Retinotopy Shape Spatial discrimination Words
	Execution	Hand flexion Oculomotor saccades Saccades Somatopy
Cognition	Music Preparation Speech	Articulatory coding
	Attention	Divided Motion Selection Selective Letter recognition First letter word generation Orthography Phonology Phonological discrimination Phonological processing Primed words Recalled words Semantic association Semantic categorization Semantic generation Semantics Words Word generation
Emotion Disease Drug	Mathematics Memory	Primed words Recalled words
		Anxiety

Table A.2: BrainMap™ “Behavioral effects”. Only a part of the behavioral effects is shown as there are approximately 300 different types.

Appendix B

Derivations

B.1 Principal component analysis as constrained variance maximization

Theorem B.1.1 *With maximization of the variance of a linear project $\mathbf{y} = \mathbf{X}\mathbf{v}$ under the constraint $\mathbf{v}^\top \mathbf{v} = 1$ the parameters \mathbf{v} will be determined as the maximal eigenvectors of the covariance matrix of \mathbf{x}*

$$\hat{\mathbf{v}} = \max_{\lambda} \text{eig}(\mathbf{S}_{\mathbf{xx}}). \quad (\text{B.1})$$

Proof B.1.1

$$E_{\text{PCA}} = \mathbf{V}[\mathbf{y}] \quad \wedge \quad \mathbf{v}^\top \mathbf{v} = 1 \quad (\text{B.2})$$

$$= N^{-1} \mathbf{y}^\top \mathbf{H} \mathbf{y} \quad \wedge \quad \mathbf{v}^\top \mathbf{v} = 1 \quad (\text{B.3})$$

$$= N^{-1} (\mathbf{X}\mathbf{v})^\top \mathbf{H} \mathbf{X} \mathbf{v} \quad \wedge \quad \mathbf{v}^\top \mathbf{v} = 1 \quad (\text{B.4})$$

$$= N^{-1} \mathbf{v}^\top \mathbf{X}^\top \mathbf{H} \mathbf{X} \mathbf{v} \quad \wedge \quad \mathbf{v}^\top \mathbf{v} = 1 \quad (\text{B.5})$$

$$= \mathbf{v}^\top \mathbf{S}_{\mathbf{xx}} \mathbf{v} \quad \wedge \quad \mathbf{v}^\top \mathbf{v} = 1 \quad (\text{B.6})$$

The maximum of the first part under the constraint can be found with Mardia et al. (1979, theorem A.9.2) and is simply the maximal eigenvector, — the eigenvector corresponding to the largest eigenvalue

$$\hat{\mathbf{v}} = \max_{\lambda} \text{eig}(\mathbf{S}_{\mathbf{xx}}). \quad \blacksquare \quad (\text{B.7})$$

B.2 Ridge regression and singular value decomposition

In connection with using canonical ridge analysis on large singular matrices the ridge penalization does not have to be applied in the large space but can be confined to a smaller space, e.g., with dimension $K = \text{rank}(\mathbf{X}^\top \mathbf{Y})$. Somewhat similar derivations appear in (Kustra and Strother 2000, appendix A), (Kustra 2000, appendix F).

Lemma B.2.1 (Exponential of ridge regularized covariance for singular data matrix) *With the matrix $\mathbf{X}(N \times P)$ of rank $K = \text{rank}(\mathbf{X})$ singular value decomposed $\mathbf{U}\mathbf{L}\mathbf{V}^\top = \text{svd}[\mathbf{X}]$, β is a rational number and $k > 0$:*

$$(\mathbf{X}^\top \mathbf{X} + k\mathbf{I}_P)^\beta = \mathbf{V} \mathbf{\Lambda} \mathbf{V}^\top, \quad (\text{B.8})$$

where the diagonal matrix $\mathbf{\Lambda}$ is defined as

$$\mathbf{\Lambda} = (\mathbf{L}^2 + k\mathbf{I}_K)^\beta. \quad (\text{B.9})$$

Proof B.2.1

$$(\mathbf{X}^\top \mathbf{X} + k\mathbf{I}_P)^\beta = \left[(\mathbf{U}\mathbf{L}\mathbf{V}^\top)^\top \mathbf{U}\mathbf{L}\mathbf{V}^\top + k\mathbf{I}_P \right]^\beta \quad (\text{B.10})$$

$$= (\mathbf{V}\mathbf{L}\mathbf{U}^\top \mathbf{U}\mathbf{L}\mathbf{V}^\top + k\mathbf{I}_P)^\beta \quad (\text{B.11})$$

$$= (\mathbf{V}\mathbf{L}^2 \mathbf{V}^\top + k\mathbf{I}_P)^\beta \quad (\text{B.12})$$

$$= (\mathbf{V}\mathbf{L}^2 \mathbf{V}^\top + \mathbf{V}k\mathbf{I}_K \mathbf{V}^\top)^\beta \quad (\text{B.13})$$

$$= [\mathbf{V} (\mathbf{L}^2 + k\mathbf{I}_K) \mathbf{V}^\top]^\beta \quad (\text{B.14})$$

With (Mardia et al. 1979, corollary A.6.4.1):

$$= \mathbf{V} (\mathbf{L}^2 + k \mathbf{I}_K)^\beta \mathbf{V}^\top. \quad \blacksquare \quad (\text{B.15})$$

Lemma B.2.2 (Covariance of singular ridge regularized prewhitened matrices) *With the singular value decomposition of two matrices $\mathbf{X} = \mathbf{U}\mathbf{L}\mathbf{V}^\top$ and $\mathbf{Y} = \mathbf{C}\mathbf{D}\mathbf{E}^\top$ the ridge regularized prewhitening of the covariance between the two matrices becomes*

$$(\mathbf{X}^\top \mathbf{X} + k_x \mathbf{I}_P)^{-1/2} \mathbf{X}^\top \mathbf{Y} (\mathbf{Y}^\top \mathbf{Y} + k_y \mathbf{I}_Q)^{-1/2} = \mathbf{V} \mathbf{\Lambda}_x \mathbf{U}^\top \mathbf{C} \mathbf{\Lambda}_y \mathbf{E}^\top, \quad (\text{B.16})$$

With the diagonal matrices defined as

$$\mathbf{\Lambda}_x = (\mathbf{L}^2 + k_x \mathbf{I}_K)^{-1/2} \mathbf{L} \quad (\text{B.17})$$

$$\mathbf{\Lambda}_y = (\mathbf{D}^2 + k_y \mathbf{I}_L)^{-1/2} \mathbf{D}, \quad (\text{B.18})$$

where $K = \text{rank}(\mathbf{X})$ and $L = \text{rank}(\mathbf{Y})$.

Proof B.2.2 The proof is shown for the first part of equation B.16 involving \mathbf{X} . The other part follows from a symmetry argument.

$$(\mathbf{X}^\top \mathbf{X} + k_x \mathbf{I}_P)^{-1/2} \mathbf{X}^\top = (\mathbf{X}^\top \mathbf{X} + k_x \mathbf{I}_P)^{-1/2} (\mathbf{U}\mathbf{L}\mathbf{V}^\top)^\top \quad (\text{B.19})$$

$$= (\mathbf{X}^\top \mathbf{X} + k_x \mathbf{I}_P)^{-1/2} \mathbf{V} \mathbf{L} \mathbf{U}^\top \quad (\text{B.20})$$

By using lemma B.2.1

$$= \mathbf{V} (\mathbf{L}^2 + k_x \mathbf{I}_K)^{-1/2} \mathbf{V}^\top \mathbf{V} \mathbf{L} \mathbf{U}^\top \quad (\text{B.21})$$

$$= \mathbf{V} (\mathbf{L}^2 + k_x \mathbf{I}_K)^{-1/2} \mathbf{L} \mathbf{U}^\top \quad (\text{B.22})$$

$$= \mathbf{V} \mathbf{\Lambda}_x \mathbf{U}^\top \quad (\text{B.23})$$

With the diagonal matrix defined as

$$\mathbf{\Lambda}_x = (\mathbf{L}^2 + k_x \mathbf{I}_K)^{-1/2} \mathbf{L} \quad (\text{B.24})$$

$$= \text{diag} \left(\frac{l_1}{\sqrt{l_1^2 + k_x}}, \dots, \frac{l_K}{\sqrt{l_K^2 + k_x}} \right), \quad (\text{B.25})$$

where l_k is the k 'th diagonal element. \blacksquare

Theorem B.2.1 (Canonical correlation vectors from canonical ridge analysis of singular matrices) *If the singular value decomposition of two matrices $\mathbf{X}(N \times P)$ and $\mathbf{Y}(N \times Q)$ are $\mathbf{X} = \mathbf{U}\mathbf{L}\mathbf{V}^\top$ and $\mathbf{Y} = \mathbf{C}\mathbf{D}\mathbf{E}^\top$ the canonical correlation vectors $\mathbf{A}(P \times K)$ and $\mathbf{B}(Q \times K)$ and the diagonal matrix \mathbf{R} with canonical correlation coefficients are found as*

$$\mathbf{A} = \mathbf{V} \check{\mathbf{\Lambda}}_x \mathbf{U}^\top \check{\mathbf{P}} \quad (\text{B.26})$$

$$\mathbf{B} = \mathbf{E} \check{\mathbf{\Lambda}}_y \mathbf{C}^\top \check{\mathbf{Q}} \quad (\text{B.27})$$

$$\mathbf{R} = \mathbf{\Psi}^{1/2} = \mathbf{\Upsilon}^{1/2}, \quad (\text{B.28})$$

where $\check{\mathbf{P}}(N \times K)$ and $\check{\mathbf{Q}}(N \times K)$ are scaled right eigenvectors and the diagonal matrices $\mathbf{\Psi}(K \times K)$ and $\mathbf{\Upsilon}(K \times K)$ contain K eigenvalues each from a symmetric $K \times K$ eigenvalue decomposition

$$\mathbf{P} \mathbf{\Psi} \mathbf{P}^\top = \text{reig}(\mathbf{C} \mathbf{\Lambda}_y^2 \mathbf{C}^\top \mathbf{U} \mathbf{\Lambda}_x^2 \mathbf{U}^\top) \quad (\text{B.29})$$

$$\mathbf{Q} \mathbf{\Upsilon} \mathbf{Q}^\top = \text{reig}(\mathbf{U} \mathbf{\Lambda}_x^2 \mathbf{U}^\top \mathbf{C} \mathbf{\Lambda}_y^2 \mathbf{C}^\top) \quad (\text{B.30})$$

$$\check{\mathbf{P}} = \mathbf{P} \left\{ \text{diag}[\text{diag}(\mathbf{P}^\top \mathbf{U} \mathbf{\Lambda}_x^2 \mathbf{U}^\top \mathbf{P})] \right\}^{-1/2} \quad (\text{B.31})$$

$$\check{\mathbf{Q}} = \mathbf{Q} \left\{ \text{diag}[\text{diag}(\mathbf{Q}^\top \mathbf{C} \mathbf{\Lambda}_y^2 \mathbf{C}^\top \mathbf{Q})] \right\}^{-1/2} \quad (\text{B.32})$$

$$\check{\mathbf{\Lambda}}_x = (\mathbf{L}^2 + k_x \mathbf{I}_K)^{-1} \mathbf{L} \quad (\text{B.33})$$

$$\check{\mathbf{\Lambda}}_y = (\mathbf{D}^2 + k_y \mathbf{I}_L)^{-1} \mathbf{D} \quad (\text{B.34})$$

$$\mathbf{\Lambda}_x = (\mathbf{L}^2 + k_x \mathbf{I}_K)^{-1/2} \mathbf{L} \quad (\text{B.35})$$

$$\mathbf{\Lambda}_y = (\mathbf{D}^2 + k_y \mathbf{I}_L)^{-1/2} \mathbf{D}, \quad (\text{B.36})$$

where \mathbf{B} may also be computed as

$$\mathbf{B} = \mathbf{E} \check{\mathbf{\Lambda}}_y \mathbf{C}^T \mathbf{U} \mathbf{\Lambda}_x^2 \mathbf{U}^T \check{\mathbf{P}} \quad (\text{B.37})$$

Proof B.2.3 With an abusive notation the canonical correlation vectors and coefficients can be written as

$$\mathbf{A} \mathbf{R} \mathbf{B}^T = (\mathbf{X}^T \mathbf{X} + k_x \mathbf{I}_P)^{-1/2} \text{svd} \left[(\mathbf{X}^T \mathbf{X} + k_x \mathbf{I}_P)^{-1/2} \mathbf{X}^T \mathbf{Y} (\mathbf{Y}^T \mathbf{Y} + k_y \mathbf{I}_Q)^{-1/2} \right] (\mathbf{Y}^T \mathbf{Y} + k_y \mathbf{I}_Q)^{-1/2} \quad (\text{B.38})$$

With lemma B.2.2

$$= (\mathbf{X}^T \mathbf{X} + k_x \mathbf{I}_P)^{-1/2} \text{svd} \underbrace{[\mathbf{V} \mathbf{\Lambda}_x \mathbf{U}^T \mathbf{C} \mathbf{\Lambda}_y \mathbf{E}^T]}_{\mathbf{K}} (\mathbf{Y}^T \mathbf{Y} + k_y \mathbf{I}_Q)^{-1/2} \quad (\text{B.39})$$

The singular vector and values of the matrix $\mathbf{K} (P \times Q)$ are now found via eigenvalue decomposition of the outer product matrix

$$\mathbf{K} \mathbf{K}^T = \mathbf{V} \mathbf{\Lambda}_x \mathbf{U}^T \mathbf{C} \mathbf{\Lambda}_y \mathbf{E}^T \mathbf{E} \mathbf{\Lambda}_y \mathbf{C}^T \mathbf{U} \mathbf{\Lambda}_x \mathbf{V}^T \quad (\text{B.40})$$

$$= \underbrace{\mathbf{V} \mathbf{\Lambda}_x \mathbf{U}^T}_{\mathbf{F} (P \times N)} \underbrace{\mathbf{C} \mathbf{\Lambda}_y^2 \mathbf{C}^T \mathbf{U} \mathbf{\Lambda}_x \mathbf{V}^T}_{\mathbf{G} (N \times P)} \quad (\text{B.41})$$

The matrices \mathbf{F} and \mathbf{G} are permuted

$$\mathbf{G} \mathbf{F} = \mathbf{C} \mathbf{\Lambda}_y^2 \mathbf{C}^T \mathbf{U} \mathbf{\Lambda}_x \mathbf{V}^T \mathbf{V} \mathbf{\Lambda}_x \mathbf{U}^T \quad (\text{B.42})$$

$$= \mathbf{C} \mathbf{\Lambda}_y^2 \mathbf{C}^T \mathbf{U} \mathbf{\Lambda}_x^2 \mathbf{U}^T, \quad (\text{B.43})$$

and the matrices $\mathbf{G} \mathbf{F}$ and $\mathbf{F} \mathbf{G}$ are right eigenvalue decomposed

$$\mathbf{P} \mathbf{\Psi} \mathbf{P}^T = \text{reig} (\mathbf{G} \mathbf{F}) \quad (\text{B.44})$$

$$\mathbf{\Pi} \mathbf{\Pi}^T = \text{reig} (\mathbf{F} \mathbf{G}) \quad (\text{B.45})$$

The two eigenvalue decomposition are related through (Mardia et al. 1979, theorem A.6.2), i.e. the non-zero eigenvalues are equal $\mathbf{\Psi} = \mathbf{\Gamma} = \mathbf{R}^2$ (the square of the singular values) and the eigenvectors are found as

$$\mathbf{\Pi} \underset{\text{columns}}{\propto} \tilde{\mathbf{\Pi}} = \mathbf{F} \mathbf{P} \quad (\text{B.46})$$

$$= \mathbf{V} \mathbf{\Lambda}_x \mathbf{U}^T \mathbf{P} \quad (\text{B.47})$$

The eigenvector matrix $\tilde{\mathbf{\Pi}}$ can be scaled so it contains standardized eigenvectors

$$\mathbf{\Pi} = \tilde{\mathbf{\Pi}} \left\{ \text{diag} \left[\text{diag} \left(\tilde{\mathbf{\Pi}}^T \tilde{\mathbf{\Pi}} \right) \right] \right\}^{-1/2} \quad (\text{B.48})$$

$$= \mathbf{V} \mathbf{\Lambda}_x \mathbf{U}^T \mathbf{P} \left\{ \text{diag} \left[\text{diag} \left(\mathbf{P}^T \mathbf{U} \mathbf{\Lambda}_x \mathbf{V}^T \mathbf{V} \mathbf{\Lambda}_x \mathbf{U}^T \mathbf{P} \right) \right] \right\}^{-1/2} \quad (\text{B.49})$$

$$= \mathbf{V} \mathbf{\Lambda}_x \mathbf{U}^T \mathbf{P} \left\{ \text{diag} \left[\text{diag} \left(\mathbf{P}^T \mathbf{U} \mathbf{\Lambda}_x^2 \mathbf{U}^T \mathbf{P} \right) \right] \right\}^{-1/2}. \quad (\text{B.50})$$

To arrive at the matrix \mathbf{A} canonical correlation vector the square root of the ridge regularized covariance should be applied (Mardia et al. 1979, definition 10.2.8) to the eigenvectors

$$\mathbf{A} = (\mathbf{X}^T \mathbf{X} + k_x \mathbf{I}_P)^{-1/2} \mathbf{\Pi} \quad (\text{B.51})$$

With lemma B.2.1 and with $\tilde{\mathbf{\Lambda}}_x = (\mathbf{L}^2 + k_x \mathbf{I}_K)^{-1/2}$

$$= \mathbf{V} \tilde{\mathbf{\Lambda}}_x \mathbf{V}^T \mathbf{\Pi} \quad (\text{B.52})$$

and inserting equation B.50

$$= \mathbf{V} \tilde{\mathbf{\Lambda}}_x \mathbf{V}^T \mathbf{V} \mathbf{\Lambda}_x \mathbf{U}^T \mathbf{P} \left\{ \text{diag} \left[\text{diag} \left(\mathbf{P}^T \mathbf{U} \mathbf{\Lambda}_x^2 \mathbf{U}^T \mathbf{P} \right) \right] \right\}^{-1/2} \quad (\text{B.53})$$

with $\check{\mathbf{\Lambda}}_x = \tilde{\mathbf{\Lambda}}_x \mathbf{\Lambda}_x$

$$= \mathbf{V} \check{\mathbf{\Lambda}}_x \mathbf{U}^T \mathbf{P} \left\{ \text{diag} \left[\text{diag} \left(\mathbf{P}^T \mathbf{U} \mathbf{\Lambda}_x^2 \mathbf{U}^T \mathbf{P} \right) \right] \right\}^{-1/2} \quad (\text{B.54})$$

and the definition $\check{\mathbf{P}} = \mathbf{P} \left\{ \text{diag} \left[\text{diag} \left(\mathbf{P}^T \mathbf{U} \mathbf{\Lambda}_x^2 \mathbf{U}^T \mathbf{P} \right) \right] \right\}^{-1/2}$

$$= \mathbf{V} \check{\mathbf{\Lambda}}_x \mathbf{U}^T \check{\mathbf{P}} \quad (\text{B.55})$$

The result for \mathbf{B} follows either from symmetric derivation or

$$\mathbf{K}^T \mathbf{K} = \underbrace{\mathbf{E} \mathbf{\Lambda}_y \mathbf{C}^T \mathbf{U} \mathbf{\Lambda}_x^2 \mathbf{U}^T}_{\mathbf{M}(Q \times N)} \underbrace{\mathbf{C} \mathbf{\Lambda}_y \mathbf{E}^T}_{\mathbf{N}(N \times Q)} \quad (\text{B.56})$$

\mathbf{NM} is the same matrix as \mathbf{GF} and thus the non-standardized eigenvectors of \mathbf{MN} , i.e., $\mathbf{K}^T \mathbf{K}$ can be found as

$$\mathbf{\Xi} \mathbf{\Omega} \mathbf{\Xi}^T = \text{reig}(\mathbf{MN}) \quad (\text{B.57})$$

$$\mathbf{\Xi} = \mathbf{M} \mathbf{P} \quad (\text{B.58})$$

$$= \mathbf{E} \mathbf{\Lambda}_y \mathbf{C}^T \mathbf{U} \mathbf{\Lambda}_x^2 \mathbf{U}^T \mathbf{P} \quad (\text{B.59})$$

$$\tilde{\mathbf{B}} = (\mathbf{Y}^T \mathbf{Y} + k_y \mathbf{I}_Q)^{-1/2} \mathbf{\Xi} \quad (\text{B.60})$$

$$= \mathbf{E} \tilde{\mathbf{\Lambda}}_y \mathbf{E}^T \mathbf{E} \mathbf{\Lambda}_y \mathbf{C}^T \mathbf{U} \mathbf{\Lambda}_x^2 \mathbf{U}^T \mathbf{P} \quad (\text{B.61})$$

$$= \mathbf{E} \check{\mathbf{\Lambda}}_y \mathbf{C}^T \mathbf{U} \mathbf{\Lambda}_x^2 \mathbf{U}^T \mathbf{P} \quad \blacksquare \quad (\text{B.62})$$

B.3 Mutual information

Find the mutual information between two multivariate Gaussian variables:

$$I(\mathbf{x}; \mathbf{y}) = h(\mathbf{x}) + h(\mathbf{y}) - h(\mathbf{x}, \mathbf{y}) \quad (\text{B.63})$$

Plug-in the entropy for the multivariate Gaussian distribution:

$$= \frac{1}{2} \ln [(2\pi e)^P |\mathbf{\Sigma}_{\mathbf{xx}}|] + \frac{1}{2} \ln [(2\pi e)^Q |\mathbf{\Sigma}_{\mathbf{yy}}|] - \frac{1}{2} \ln \left[(2\pi e)^{(P+Q)} \begin{vmatrix} \mathbf{\Sigma}_{\mathbf{xx}} & \mathbf{\Sigma}_{\mathbf{xy}} \\ \mathbf{\Sigma}_{\mathbf{yx}} & \mathbf{\Sigma}_{\mathbf{yy}} \end{vmatrix} \right] \quad (\text{B.64})$$

$$= \frac{1}{2} \ln \left[\frac{(2\pi e)^P |\mathbf{\Sigma}_{\mathbf{xx}}| (2\pi e)^Q |\mathbf{\Sigma}_{\mathbf{yy}}|}{(2\pi e)^{(P+Q)} \begin{vmatrix} \mathbf{\Sigma}_{\mathbf{xx}} & \mathbf{\Sigma}_{\mathbf{xy}} \\ \mathbf{\Sigma}_{\mathbf{yx}} & \mathbf{\Sigma}_{\mathbf{yy}} \end{vmatrix}} \right] \quad (\text{B.65})$$

$$= \frac{1}{2} \ln \frac{|\mathbf{\Sigma}_{\mathbf{xx}}| |\mathbf{\Sigma}_{\mathbf{yy}}|}{\begin{vmatrix} \mathbf{\Sigma}_{\mathbf{xx}} & \mathbf{\Sigma}_{\mathbf{xy}} \\ \mathbf{\Sigma}_{\mathbf{yx}} & \mathbf{\Sigma}_{\mathbf{yy}} \end{vmatrix}} \quad (\text{B.66})$$

Mardia, Kent, and Bibby (1979, equation A.2.3j):

$$= \frac{1}{2} \ln \frac{|\mathbf{\Sigma}_{\mathbf{xx}}| |\mathbf{\Sigma}_{\mathbf{yy}}|}{|\mathbf{\Sigma}_{\mathbf{xx}}| |\mathbf{\Sigma}_{\mathbf{yy}} - \mathbf{\Sigma}_{\mathbf{yx}} \mathbf{\Sigma}_{\mathbf{xx}}^{-1} \mathbf{\Sigma}_{\mathbf{xy}}|} \quad (\text{B.67})$$

$$= \frac{1}{2} \ln \frac{|\mathbf{\Sigma}_{\mathbf{yy}}|}{|\mathbf{\Sigma}_{\mathbf{yy}} - \mathbf{\Sigma}_{\mathbf{yx}} \mathbf{\Sigma}_{\mathbf{xx}}^{-1} \mathbf{\Sigma}_{\mathbf{xy}}|} \quad (\text{B.68})$$

Mardia, Kent, and Bibby (1979, equation A.2.3k):

$$= \frac{1}{2} \ln \frac{|\mathbf{\Sigma}_{\mathbf{yy}}|}{|\mathbf{\Sigma}_{\mathbf{yy}}| \mathbf{I}_Q + \mathbf{\Sigma}_{\mathbf{yy}}^{-1} \mathbf{\Sigma}_{\mathbf{yx}} \mathbf{\Sigma}_{\mathbf{xx}}^{-1} \mathbf{\Sigma}_{\mathbf{xy}}|} \quad (\text{B.69})$$

$$= \frac{1}{2} \ln \frac{1}{|\mathbf{I}_Q + \mathbf{\Sigma}_{\mathbf{yy}}^{-1} \mathbf{\Sigma}_{\mathbf{yx}} \mathbf{\Sigma}_{\mathbf{xx}}^{-1} \mathbf{\Sigma}_{\mathbf{xy}}|} \quad (\text{B.70})$$

$$= -\frac{1}{2} \ln |\mathbf{I}_Q + \mathbf{\Sigma}_{\mathbf{yy}}^{-1} \mathbf{\Sigma}_{\mathbf{yx}} \mathbf{\Sigma}_{\mathbf{xx}}^{-1} \mathbf{\Sigma}_{\mathbf{xy}}|. \quad (\text{B.71})$$

Which is the same as (Mardia, Kent, and Bibby 1979, equation A.2.3n):

$$= -\frac{1}{2} \ln |\mathbf{I}_P + \mathbf{\Sigma}_{\mathbf{xx}}^{-1} \mathbf{\Sigma}_{\mathbf{xy}} \mathbf{\Sigma}_{\mathbf{yy}}^{-1} \mathbf{\Sigma}_{\mathbf{yx}}|. \quad (\text{B.72})$$

B.4 The Bilinear model

B.4.1 Canonical correlation analysis

Theorem B.4.1 (CCA as part of BLM) *With the bilinear model*

$$\mathbf{X} = \mathbf{T}\mathbf{P} + \mathbf{E} \quad (\text{B.73})$$

$$\mathbf{Y} = \mathbf{T}\mathbf{Q} + \mathbf{F} \quad (\text{B.74})$$

together with the conditions

$$\hat{\mathbf{T}} = \mathbf{X}\Phi \quad (\text{B.75})$$

$$\Sigma_{ee} = \mathbf{X}^T \mathbf{X} \quad (\text{B.76})$$

$$\Sigma_{ff} = \mathbf{Y}^T \mathbf{Y}, \quad (\text{B.77})$$

the maximum likelihood estimate of Φ is the left canonical correlation vectors of \mathbf{X} and \mathbf{Y} .

Proof B.4.1 The maximum likelihood estimate of Φ is found by the solution to the generalized eigenvalue problem (Burnham, MacGregor, and Viveros 1999, equation 10)

$$(\mathbf{X}^T \mathbf{X} \Sigma_{ee}^{-1} \mathbf{X}^T \mathbf{X} + \mathbf{X}^T \mathbf{Y} \Sigma_{ff}^{-1} \mathbf{Y}^T \mathbf{X}) \Phi = \mathbf{X}^T \mathbf{X} \Phi \mathbf{D} \quad (\text{B.78})$$

Inserting equations B.76 and B.77 and substituting $\mathbf{S}_{xx} = \mathbf{X}^T \mathbf{X}$ and $\mathbf{S}_{yy} = \mathbf{Y}^T \mathbf{Y}$

$$(\mathbf{S}_{xx} \mathbf{S}_{xx}^{-1} \mathbf{S}_{xx} + \mathbf{S}_{xy} \mathbf{S}_{yy}^{-1} \mathbf{S}_{yx}) \Phi = \mathbf{S}_{xx} \Phi \mathbf{D} \quad (\text{B.79})$$

With an invertable \mathbf{S}_{xx}

$$\mathbf{S}_{xx}^{-1} (\mathbf{S}_{xx} \mathbf{S}_{xx}^{-1} \mathbf{S}_{xx} + \mathbf{S}_{xy} \mathbf{S}_{yy}^{-1} \mathbf{S}_{yx}) \Phi = \Phi \mathbf{D} \quad (\text{B.80})$$

$$(\mathbf{S}_{xx}^{-1} \mathbf{S}_{xx} \mathbf{S}_{xx}^{-1} \mathbf{S}_{xx} + \mathbf{S}_{xx}^{-1} \mathbf{S}_{xy} \mathbf{S}_{yy}^{-1} \mathbf{S}_{yx}) \Phi = \Phi \mathbf{D} \quad (\text{B.81})$$

$$(\mathbf{I}_P + \mathbf{S}_{xx}^{-1} \mathbf{S}_{xy} \mathbf{S}_{yy}^{-1} \mathbf{S}_{yx}) \Phi = \Phi \mathbf{D} \quad (\text{B.82})$$

$$\Phi + \mathbf{S}_{xx}^{-1} \mathbf{S}_{xy} \mathbf{S}_{yy}^{-1} \mathbf{S}_{yx} \Phi = \Phi \mathbf{D} \quad (\text{B.83})$$

$$\mathbf{S}_{xx}^{-1} \mathbf{S}_{xy} \mathbf{S}_{yy}^{-1} \mathbf{S}_{yx} \Phi = \Phi \mathbf{D} - \Phi \quad (\text{B.84})$$

$$\mathbf{S}_{xx}^{-1} \mathbf{S}_{xy} \mathbf{S}_{yy}^{-1} \mathbf{S}_{yx} \Phi = \Phi (\mathbf{D} - \mathbf{I}) \quad (\text{B.85})$$

With $\tilde{\mathbf{D}} = \mathbf{D} - \mathbf{I}$

$$\mathbf{S}_{xx}^{-1} \mathbf{S}_{xy} \mathbf{S}_{yy}^{-1} \mathbf{S}_{yx} \Phi = \Phi \tilde{\mathbf{D}}. \quad (\text{B.86})$$

The square root of the eigenvalues in $\tilde{\mathbf{D}}$ are the canonical correlation coefficients and Φ is identified as the left canonical correlation vectors. ■

B.5 Molgedey-Schuster ICA

The Molgedey Schuster transformation can be derived by regarding an autoregression function from (time/space) shift $t = 0$ to $t = \tau$ with the regression coefficients in \mathbf{W} and additive noise \mathbf{U}

$$\mathbf{X}_\tau = \mathbf{X}\mathbf{W} + \mathbf{U}. \quad (\text{B.87})$$

With the additive noise independent in both in objects and variables

$$\mathbf{V}[\mathbf{U}^V] = \mathbf{I}_P \otimes \mathbf{I}_N, \quad (\text{B.88})$$

the parameters can be estimated with maximum likelihood (ordinary least squares solution)

$$\hat{\mathbf{W}} = [\mathbf{X}^T \mathbf{X}]^{-1} \mathbf{X}^T \mathbf{X}_\tau. \quad (\text{B.89})$$

Now the $\mathbf{X}(N \times P)$ are regarded as observations generated from K unknown sources $\mathbf{Z}(N \times K)$ that are mixed linearly using a mixing matrix $\mathbf{A}(K \times P)$:

$$\mathbf{X} = \mathbf{Z}\mathbf{A}. \quad (\text{B.90})$$

The sources are regarded as independent between the K variables (i.e. the K sources):

$$\mathbf{Z}^\top \mathbf{Z} = \mathbf{\Lambda}, \quad (\text{B.91})$$

where $\mathbf{\Lambda}$ is a diagonal matrix denoting the covariance. The covariance between $t = 0$ and $t = \tau$ should also be regarded as independent among the K variables of the sources:

$$\mathbf{Z}^\top \mathbf{Z}_\tau = \mathbf{\Lambda}_\tau. \quad (\text{B.92})$$

\mathbf{X} is now substituted with the sources and the mixing matrix, i.e., expand equation B.89 from equation B.90

$$\hat{\mathbf{W}} = [(\mathbf{Z}\mathbf{A})^\top \mathbf{Z}\mathbf{A}]^{-1} (\mathbf{Z}\mathbf{A})^\top \mathbf{Z}_\tau \mathbf{A} \quad (\text{B.93})$$

$$\hat{\mathbf{W}} = [\mathbf{A}^\top \mathbf{Z}^\top \mathbf{Z}\mathbf{A}]^{-1} \mathbf{A}^\top \mathbf{Z}^\top \mathbf{Z}_\tau \mathbf{A} \quad (\text{B.94})$$

$$\hat{\mathbf{W}} = [\mathbf{A}^\top \mathbf{\Lambda}\mathbf{A}]^{-1} \mathbf{A}^\top \mathbf{\Lambda}_\tau \mathbf{A} \quad (\text{B.95})$$

$$\mathbf{A}^\top \mathbf{\Lambda}\mathbf{A} \hat{\mathbf{W}} = \mathbf{A}^\top \mathbf{\Lambda}_\tau \mathbf{A} \quad (\text{B.96})$$

$$\mathbf{\Lambda}\mathbf{A} \hat{\mathbf{W}} = \mathbf{\Lambda}_\tau \mathbf{A} \quad (\text{B.97})$$

$$\mathbf{A} \hat{\mathbf{W}} = \mathbf{\Lambda}^{-1} \mathbf{\Lambda}_\tau \mathbf{A}, \quad (\text{B.98})$$

$$\hat{\mathbf{W}} = \mathbf{A}^{-1} \mathbf{\Lambda}^{-1} \mathbf{\Lambda}_\tau \mathbf{A}, \quad (\text{B.99})$$

With a diagonal matrix $\mathbf{L} = \mathbf{\Lambda}^{-1} \mathbf{\Lambda}_\tau$ regarded as the eigenvalues the right side terms can be regarded as the left eigenvalue decomposition of $\hat{\mathbf{W}}$:

$$\text{leig}[\hat{\mathbf{W}}] = \mathbf{A}^{-1} \mathbf{L} \mathbf{A}. \quad (\text{B.100})$$

Thus the mixing matrix \mathbf{A} is identified as the left eigenvector of $\hat{\mathbf{W}}$.

Theorem B.5.1 (Molgedey-Schuster ICA with SVD) *With the singular value decompositions of the original matrix $\mathbf{U}\mathbf{L}\mathbf{V}^\top = \text{svd}(\mathbf{X})$ and the shifted matrix $\mathbf{U}_\tau \mathbf{L}\mathbf{V}^\top = \text{svd}(\mathbf{X}_\tau)$ the source \mathbf{Z} and mixing matrix \mathbf{A} from a symmetrical Molgedey-Schuster independent component analysis is found as*

$$\mathbf{A} = \tilde{\mathbf{A}} \mathbf{L} \mathbf{V}^\top. \quad (\text{B.101})$$

$$\mathbf{Z} = \mathbf{U} \tilde{\mathbf{A}}^{-1}. \quad (\text{B.102})$$

where $\tilde{\mathbf{A}}$ is the left eigenvectors

$$\tilde{\mathbf{A}}^{-1} \tilde{\mathbf{\Gamma}} \tilde{\mathbf{A}} = \frac{1}{2} \text{leig}(\mathbf{U}^\top \mathbf{U}_\tau + \mathbf{U}_\tau^\top \mathbf{U}). \quad (\text{B.103})$$

A derivation appears in (Hansen and Larsen 1998) differing with respect to the one shown below in the scaling between the source and the mixing matrix.

Proof B.5.1 Symmetrized Molgedey-Schuster ICA finds the left eigenvectors of the following matrix

$$\mathbf{W} = \frac{1}{2} (\mathbf{X}^\top \mathbf{X})^{-1} (\mathbf{X}^\top \mathbf{X}_\tau + \mathbf{X}_\tau^\top \mathbf{X}) \quad (\text{B.104})$$

The data matrices are substituted with the singular value decompositions $\mathbf{U}\mathbf{L}\mathbf{V}^\top = \text{svd}(\mathbf{X})$ and $\mathbf{U}_\tau \mathbf{L}\mathbf{V}^\top = \text{svd}(\mathbf{X}_\tau)$

$$\mathbf{W} = \frac{1}{2} \mathbf{V} \mathbf{L}^{-2} \mathbf{V}^\top (\mathbf{V} \mathbf{L} \mathbf{U}^\top \mathbf{U}_\tau^\top \mathbf{L} \mathbf{V}^\top + \mathbf{V} \mathbf{L} \mathbf{U}_\tau^\top \mathbf{U} \mathbf{L} \mathbf{V}^\top) \quad (\text{B.105})$$

$$\mathbf{W} = \frac{1}{2} \mathbf{V} \mathbf{L}^{-2} (\mathbf{L} \mathbf{U}^\top \mathbf{U}_\tau \mathbf{L} \mathbf{V}^\top + \mathbf{L} \mathbf{U}_\tau^\top \mathbf{U} \mathbf{L} \mathbf{V}^\top) \quad (\text{B.106})$$

$$\mathbf{W} = \frac{1}{2} \underbrace{\mathbf{V} \mathbf{L}^{-1}}_{\mathbf{G}} \underbrace{(\mathbf{U}^\top \mathbf{U}_\tau + \mathbf{U}_\tau^\top \mathbf{U})}_{\mathbf{F}} \mathbf{L} \mathbf{V}^\top \quad (\text{B.107})$$

$$\tilde{\mathbf{W}} = \frac{1}{2} \mathbf{F} \mathbf{G} = \frac{1}{2} (\mathbf{U}^\top \mathbf{U}_\tau + \mathbf{U}_\tau^\top \mathbf{U}) \underbrace{\mathbf{L} \mathbf{V}^\top \mathbf{V} \mathbf{L}^{-1}}_{\mathbf{I}} \quad (\text{B.108})$$

$$\tilde{\mathbf{A}}^{-1} \tilde{\mathbf{\Gamma}} \tilde{\mathbf{A}} = \frac{1}{2} \text{leig}(\mathbf{U}^\top \mathbf{U}_\tau + \mathbf{U}_\tau^\top \mathbf{U}). \quad (\text{B.109})$$

With (Mardia et al. 1979, theorem A.6.2) modified for left eigenvalue problems the original mixing matrix \mathbf{A} is found as

$$\mathbf{A} = \tilde{\mathbf{A}}\mathbf{L}\mathbf{V}^T. \quad (\text{B.110})$$

The sources \mathbf{Z} are found as

$$\mathbf{Z}\mathbf{A} = \mathbf{X} \quad (\text{B.111})$$

$$\mathbf{Z}\tilde{\mathbf{A}}\mathbf{L}\mathbf{V}^T = \mathbf{U}\mathbf{L}\mathbf{V}^T \quad (\text{B.112})$$

$$\mathbf{Z}\tilde{\mathbf{A}} = \mathbf{U} \quad (\text{B.113})$$

$$\mathbf{Z} = \mathbf{U}\tilde{\mathbf{A}}^{-1}. \quad \blacksquare \quad (\text{B.114})$$

B.6 Distance between two vectors

Two univariate Gaussian distributions (or rather just functions) are considered. These should have the same variance and separated by Δ , so one may be set with the mean in $\mu_1 = 0$ and the other in $\mu_2 = \Delta$.

B.6.1 The product of two Gaussian distributions

The sum of the products over the variable x (the inner product):

$$\Gamma = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}x^2\right] \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(x-\Delta)^2\right] dx \quad (\text{B.115})$$

$$= \frac{1}{2\pi\sigma^2} \int_{-\infty}^{\infty} \exp\left[-\frac{1}{2\sigma^2}(2x^2 - 2\Delta x + \Delta^2)\right] dx \quad (\text{B.116})$$

$$= \frac{1}{2\pi\sigma^2} \int_{-\infty}^{\infty} \exp\left[-\left(\frac{1}{\sigma^2}x^2 - \frac{\Delta}{\sigma^2}x + \frac{\Delta^2}{2\sigma^2}\right)\right] dx \quad (\text{B.117})$$

This is a definite integral involving an exponential function which is listed by Spiegel (1968, formula 15.75)

$$= \frac{1}{2\pi\sigma^2} \sqrt{\frac{\pi}{\frac{1}{\sigma^2}}} \exp\left[\left(\left[\frac{\Delta}{\sigma^2}\right]^2 - 4\frac{1}{\sigma^2}\frac{\Delta^2}{2\sigma^2}\right) / \left(4\frac{1}{\sigma^2}\right)\right] \quad (\text{B.118})$$

$$= \frac{\sigma}{2\sqrt{\pi}\sigma^2} \exp\left[\left(\frac{\Delta^2}{\sigma^4} - \frac{4\Delta^2}{2\sigma^4}\right) / \frac{4}{\sigma^2}\right] \quad (\text{B.119})$$

$$= \frac{1}{\sqrt{4\pi}\sigma^2} \exp\left[-\frac{\Delta^2}{4\sigma^2}\right]. \quad (\text{B.120})$$

B.6.2 Kullback-Leibler distance

Kullback-Leibler distance with p as the reference:

$$K(p||q) = - \int_{-\infty}^{\infty} p(x) \ln \frac{q(x)}{p(x)} dx \quad (\text{B.121})$$

Insert the Gaussian distributions

$$= - \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}x^2\right) \ln \left[\frac{\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}[x-\Delta]^2\right)}{\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}x^2\right)} \right] dx \quad (\text{B.122})$$

$$= - \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2\sigma^2}x^2\right) \left[-\frac{1}{2\sigma^2}[x-\Delta]^2 + \frac{1}{2\sigma^2}x^2 \right] dx \quad (\text{B.123})$$

$$= \frac{1}{\sqrt{8\pi\sigma^6}} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2\sigma^2}x^2\right) [(x-\Delta)^2 - x^2] dx \quad (\text{B.124})$$

$$= \frac{1}{\sqrt{8\pi\sigma^6}} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2\sigma^2}x^2\right) [-2\Delta x + \Delta^2] dx \quad (\text{B.125})$$

$$= \frac{1}{\sqrt{8\pi\sigma^6}} \left[-2\Delta \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2\sigma^2}x^2\right) x dx + \Delta^2 \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2\sigma^2}x^2\right) dx \right] \quad (\text{B.126})$$

The function in the first integral is odd. The second integral is listed by Spiegel (1968, formula 15.72) and is just an ordinary unnormalized Gaussian distribution:

$$= \frac{1}{\sqrt{8\pi\sigma^6}} \left[0 + \Delta^2 \sqrt{2\pi\sigma^2} \right] \quad (\text{B.127})$$

$$= \frac{\Delta^2}{2\sigma^2}. \quad (\text{B.128})$$

Appendix C

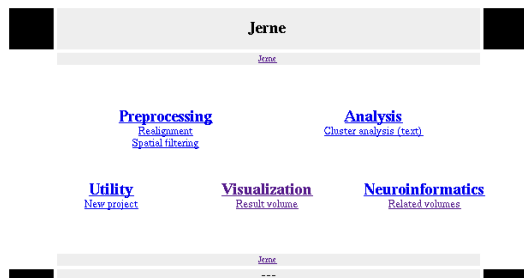
Example of web-services

Many of the analysis and visualization methods described can be extended to Internet services. Figure C.1 shows simple examples with cluster analysis and visualization of a result volume: From the main window in figure C.1(a) it is possible to get to either the cluster analysis upload window, figure C.1(b), or the result volume visualization figure C.1(e).

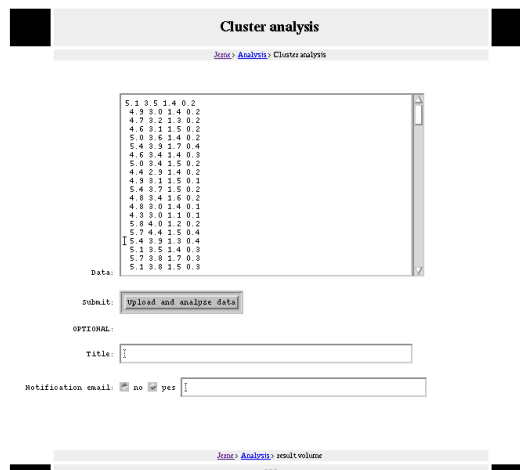
Data uploading is complex for functional neuroimaging: Apart from the very large sizes data is often distributed over several files and the form interface associated with HTML does not allow multiple file select or selection of entire directories. Furthermore, auxiliary information apart from the image data might be needed, e.g., design matrices. For small-sized data sets in ASCII representation the multi-line text HTML element can be used. This is the case for the web-page displayed in figure C.1(b). The particular data set entered is the small Iris data set (Fisher 1936; Mardia et al. 1979, table 1.2.2). A small Perl CGI script is used on the server-side to parse the data, construct a Matlab script and output the HTML web-page displayed in figure C.1(c). A Matlab job is then spawned with the constructed Matlab script — in this case the Matlab job runs the K-means clustering algorithm from the Lyngby toolbox. After a small amount of processing time the Matlab script produces plots and result files that are added to a temporary directory on the web-server accessible to the user that uploaded the file. The files are already linked to by the HTML web-page constructed by the Perl script. Figure C.1(d) shows a plot of this simple cluster analysis.

When binary files are to be handled the file upload facility in HTTP can be used, see, e.g., (Guelich, Gundavaram, and Birznieks 2000, page 97+). Figure C.1(e) shows a screenshot of a web-page involved in this procedure where a volume in the ANALYZE format (which consists of a data file and a header file) can be uploaded. For the particular visualization displayed in figure C.1(f) an example data sets from the SPM distribution shared by Paul Kinahan and Douglas C. Noll was used, see <http://www.fil.ion.ucl.ac.uk/spm/data/>: A 12 scan single subject PET study with a left hand finger opposition task, — which is part of a larger study (Kinahan and Noll 1999). Thus the activation in the left side of the brain is very likely the primary motor area corresponding to the left hand. In this case the output is to PostScript and JPEG image files as well as Matlab “fig” files, but it would be trivial to add VRML support. This last example shows that it is possible to make web-services for analysis and visualization of volumes.

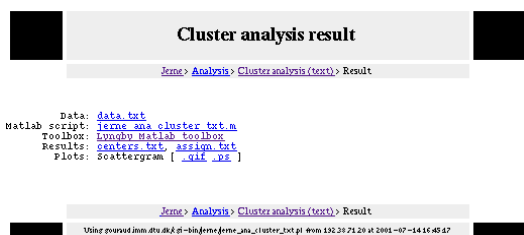
The form interface in HTML does not allow particular advanced user interactions, and the file transfer of large data set can be tedious. However, web-services can have advantages: Centralized web-services might have more computational power, so that the preprocessing and analysis are performed quickly once the data is uploaded. Motion correction requires comparable little interactive setup and a large amount of computer time. Furthermore, it is easy to parallelize and should thus be able to run easily on a (Beowulf) cluster of computers. Centralized web-services might also be able to implement new algorithms faster without the need of the user to update his/her software. Lastly, analyses that the user would not normal invoke, e.g., diagnostic, can be run in the “background”.



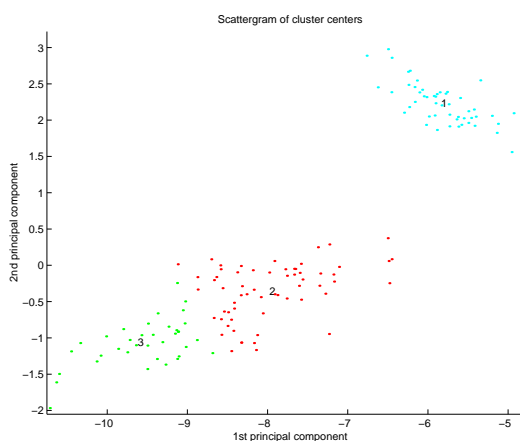
(a) Main window



(b) Cluster upload window



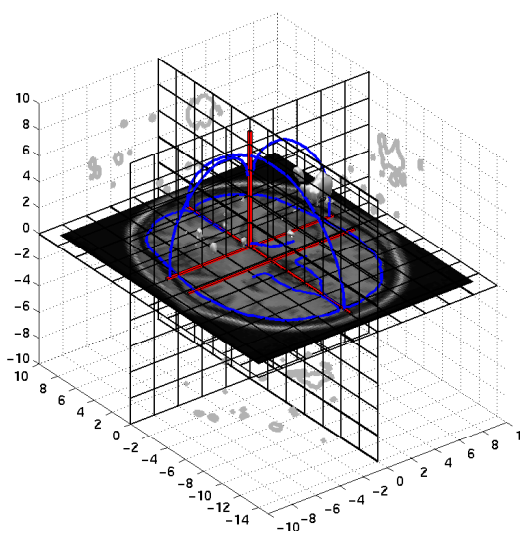
(c) Clustering result window



(d) A cluster result plot



(e) Visualization window



(f) A volume visualization result

Figure C.1: Web-services: Screenshots from pages with web-services.

Appendix D

Acknowledgment

- The thesis has been supervised by *Lars Kai Hansen* and *Jan Larsen*. Lars Kai Hansen has provided extensive support.
- I have been working together with researchers in Minneapolis at the VA Medical Center, PET Imaging Service, Center for Magnetic Resonance Research and Division of Health Computer Sciences and visited them several times. Particularly, I have had many interesting discussions with *Stephen C. Strother* and *Archana Purushotham*.
- During my Ph.D. study I spend half a year at the NEC Research Institute with the group of *C. Lee Giles*. During this time I was fortunate to be coauthor on two sets of papers primarily conceived and made by *Steve Lawrence* and *David Pennock*. One considering the persistence of URLs in scientific literature (Lawrence et al. 2001; Lawrence et al. 2000), the other considering the predictive power of artificial markets implemented as web games (Pennock, Lawrence, Giles, and Nielsen 2001c; Pennock, Lawrence, Giles, and Nielsen 2001a).
- The Lyngby Matlab toolbox has been developed in collaboration with and includes code by *Peter Toft*, *Matthew G. Liptrot*, *Lars Kai Hansen*, *Pedro Højen-Sørensen* and *Carsten Helbo*.
- During the last months of my Ph.D. I joined an informal group discussing and working with event related fMRI. *Daniela Balslev*, *Søren Kyllingsbæk*, *Matthew G. Liptrot*, *Torben Lund* and *Markus Nowak* succeeded to acquire efMRI scans.
- Many post docs, Ph.d.- and master-students in our group at the department have been helpful and influential in different aspects, particularly *Ulrik Kjems*, *Thomas Kolenda* and *Carl Edward Rasmussen*. *Cyril Goutte* “invented” the smooth FIR model and completed our group’s first description of clustering of functional neuroimages.
- *Egill Rostrup* provided fMRI scans used in the analysis chapter. Other data was provided from Minneapolis and Massachusetts General Hospital. *Ian Law* provided some volumes used in the visualizations.
- The thesis is support by the Danish Research Councils through *THOR Center for Neuroinformatics* headed by *Lars Kai Hansen*. Furthermore, I have been employed in another Danish project headed by *Olaf B. Paulson* and in the American Human Brain Project through *International Neuroimaging Consortium (INC)* headed by *David A. Rottenberg*.

Appendix E

Articles and abstracts

I have been author or listed as co-author on several published papers. Boldface entries are reprinted on the following pages.

- Time-series modeling of fMRI data
 - In **(Nielsen, Hansen, Toft, Goutte, Mørch, Svarer, Savoy, Rosen, Rostrup, and Born 1997, page 159)** the FIR model was compared to a fitted Gamma density model for the analysis of an fMRI time-series data set.
 - In **(Nielsen, Goutte, and Hansen 1999, page 160)** an artificial neural network was used as a nonlinear model for fMRI time-series.
 - **(Goutte, Nielsen, and Hansen 2000, pages 161–174)** introduced the smooth FIR with Bayesian modeling. (Nielsen, Goutte, and Hansen 2001a) is a short abstract of this work.
- Clustering of functional neuroimages
 - **(Goutte, Toft, Rostrup, Nielsen, and Hansen 1999b, pages 175–187)** introduced K-means clustering on the cross-correlation function and also examined hierarchical clustering of fMRI data. (Toft, Hansen, Nielsen, Strother, Lange, Mørch, Svarer, Paulson, Savoy, Rosen, Rostrup, and Born 1997) is an abstract of this work.
 - Feature clustering in connection with fMRI data was described in (Goutte, Nielsen, Svarer, Rostrup, and Hansen 1998).
 - K-means clustering was applied on a PET data set by Balslev, Law, Frutiger, Sidtis, Nielsen, Christiansen, Strother, Svarer, Rottenberg, Paulson, and Hansen (2000) and an article of this work is under review (Balslev, Nielsen, Frutiger, Sidtis, Christiansen, Svarer, Strother, Rottenberg, Hansen, Paulson, and Law 2001).
 - K-means clustering was also applied on single trial fMRI in (Purushotham, Nielsen, Hansen, and Kim 1999) and (Purushotham, Nielsen, Hansen, and Kim 2000).
- Other multivariate analysis of functional neuroimages
 - Canonical ridge analysis for functional neuroimages was introduced in **(Nielsen, Hansen, and Strother 1998, page 188)**.
 - Probabilistic principal component analysis was examined in **(Hansen, Larsen, Nielsen, Strother, Rostrup, Savoy, Svarer, and Paulson 1999a, pages 189–199)** with abstract (Hansen, Nielsen, Toft, Strother, Lange, Mørch, Svarer, Paulson, Savoy, Rosen, Rostrup, and Born 1997).
 - Nonlinear multivariate analysis of functional neuroimages by artificial neural networks was described in (Hansen, Mørch, and Nielsen 1998).
- Comparison of functional neuroimage analysis.
 - Nine different analysis methods were compared in **(Lange, Strother, Anderson, Nielsen, Holmes, Kolenda, Savoy, and Hansen 1999, pages 200–221)** with the abstract (Lange, Hansen, Anderson, Nielsen, Savoy, Kim, and Strother 1998).
 - In (Hansen, Nielsen, Strother, and Lange 2000c) it is shown that the consensus model computed as the averaged over histogram-equalized result images from eight different analyses on the same image data outperforms each individual model.

- Visualization
 - Functional neuroimaging visualization especially with reference to the Virtual Reality Modeling Language has been described in **(Nielsen and Hansen 1997, pages 222–225)**, **(Nielsen and Hansen 1998, page 226)**, **(Nielsen and Hansen 2000a, pages 227–232)** and (Chen, Nielsen, and Hansen 1999).
- Analysis of BrainMap™ data.
 - **(Nielsen and Hansen 1999, pages 233–239)** describe functional volumes modeling with mixture models.
 - Functional volumes modeling with kernel density estimators was described in an unpublished abstract **(Nielsen and Hansen 2000b, page 240)**.
 - **(Nielsen, Hansen, and Kjems 2001b, page 242)** identified outliers in the BrainMap™ database and a longer article is submitted to *Human Brain Mapping* **(Nielsen and Hansen 2001b, pages 243–263)**.
- Text analysis
 - In (Hansen, Sigurðsson, Kolenda, Nielsen, Kjems, and Larsen 2000d) the generalizable Gaussian mixture model was used for term analysis.
 - Author cocitation analysis was described in **(Nielsen and Hansen 2001a, pages 264–265)**.
- Miscellaneous articles.
 - The Lyngby Matlab toolbox for functional neuroimaging analysis is described in three abstracts: **(Hansen, Nielsen, Toft, Liptrot, Goutte, Strother, Lange, Gade, Rottenberg, and Paulson 1999b, page 266)** (Hansen, Nielsen, Liptrot, Goutte, Strother, Lange, Gade, Rottenberg, and Paulson 2000b) and (Hansen, Nielsen, Liptrot, Strother, Lange, Gade, Rottenberg, and Paulson 2001b)
 - The predictive performance of artificial markets implemented as web-games was examined in (Pennock, Lawrence, Giles, and Nielsen 2001c; Pennock, Lawrence, Giles, and Nielsen 2001a; Pennock, Lawrence, Giles, and Nielsen 2001b), see also <http://artificialmarkets.com>.
 - Persistence of web-citations in scientific literature was documented in (Lawrence, Coetzee, Glover, Pennock, Flake, Nielsen, Krovetz, Kruger, and Giles 2001; Lawrence, Coetzee, Flake, Pennock, Krovetz, Nielsen, Kruger, and Giles 2000)

Modeling of BrainMap data

Finn Årup Nielsen and Lars Kai Hansen

Department of Mathematical Modelling
 Technical University of Denmark
 DK-2800 Lyngby, DENMARK
 (fn,lkhansen)@imm.dtu.dk

Abstract

We apply machine learning techniques in the form of Gaussian mixture models to functional brain activation data. The dataset was extracted through the WWW interface to the BrainMap™ (Research Imaging Center, University of Texas Health Science Center at San Antonio) neuroimaging database. Modeling of the joint probability structure of activation foci and other database entries (e.g. behavioral domain, modality) enables us to summarize the accumulated body of activation coordinates in the form of a 3D density and allows us to explore issues like the effect of the experimental modality on the resulting brainmap

1 Introduction

Neuroimaging experiments based on positron emission tomography (PET), or functional magnetic resonance imaging (fMRI), are accumulating vast spatio-temporal databases at a rate that calls for new innovative informatics tools. Neuroscience databases are conceptually and physically linked in complex socio-scientific networks of human relations, publications, and funding programs. The neuroinformatics challenge is to organize these networks and make them transparent for the neuroscience community [16]. An important part of neuroinformatics concerns the process of relating different functional neuroimaging studies to each other.

A functional neuroimaging study based on fMRI or PET examines the neural correlate of a mental process as it affects cerebral blood flow, for a recent review see [14]. Under the functional segregation paradigm the brain image consists of activation hot spots that each are related to a cognitive component. When discussing the results of a functional study the observed set of activation foci is compared to foci from other studies reported in the literature most often by simple visual inspection, see e.g. a meta-analysis on visual recognition [5]. Such discussions are naturally driven primarily by the informal neuroscientific insight of the researchers involved. Description of the set of foci can be given in terms of lobes or gyri (e.g. dorsolateral prefrontal cortex) or in terms of an informal set of functional areas (such as visual areas V1, V2, ...). An increasingly popular, formal and quantitative alternative is to report foci positions with reference to the Talairach system [17] — a standardized Euclidean system of reference.

The aim of this work is to explore quantitative and automatic procedures for the comparison and discussion of functional data and in this way to pave the road for objective

Figure E.75: Unpublished article (Nielsen and Hansen 1999). Page 1 of 7.

meta-analyses. We use self-optimizing machine learning algorithms to model the densities of Talairach coordinates and use VRML (virtual reality modeling language) geometric hypertext tools as interface to the learned models.

The long term goal is a tool which can assist the neuroimaging researcher in quantifying and reporting the information content of a study with respect to the accumulated body of neuroscience.

2 BrainMap

The BrainMap™ database is an extensive collection of papers containing Talairach coordinates from human brain mapping studies maintained by the Research Imaging Center, University of Texas Health Science Center at San Antonio [6, 11]. The access to the database is provided by either a graphical user interface application or a web-based interface, see <http://ric.uthscsa.edu>. The database contains bibliographic information about the paper, formal descriptions of the study (e.g. modality, behavioral domain, response type) and 3D coordinates for the functional activations reported in the Talairach system [17]. BrainMap has been used for meta-analytic modeling under the heading “functional volumes modeling” [8, 7], see also [9, 13]. In [8] the location of the activation foci of the mouth is modeled with a bounding box. The estimation of the mouth area required a manual editing of the specific foci that were included to form the volume of interest. In this contribution we will investigate global patterns of foci found under various activation paradigms and we will use machine learning models with minimal user intervention.

3 Generalizable Gaussian Mixtures

Our primary pattern recognition device will be the Gaussian mixture, see, e.g., [15] for a review. Gaussian mixture models have been used in single study functional neuroimaging before, see e.g. [4]. The Gaussian mixture density of a datavector \mathbf{x} , is defined as

$$\begin{aligned} p(\mathbf{x}|\boldsymbol{\theta}) &= \sum_{k=1}^K P(k)p(\mathbf{x}|k) \\ p(\mathbf{x}|\boldsymbol{\theta}_k) &= \frac{1}{\sqrt{|2\pi\boldsymbol{\Sigma}_k|}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k^{-1} (\mathbf{x} - \boldsymbol{\mu}_k)\right) \end{aligned} \quad (1)$$

where the component Gaussians are mixed with proportions $\sum_k P(k) = 1$, and we have defined the parameter vector $\boldsymbol{\theta} \equiv \{\boldsymbol{\Sigma}_k, \boldsymbol{\mu}_k\}$. The parameters are estimated from a set of examples $D = \{\mathbf{x}_n | n = 1, \dots, N\}$. In the pattern recognition literature mixture densities are mostly estimated by maximum likelihood (ML), using various estimate-maximize (EM) methods [15]. The (negative log-)likelihood costfunction is defined by

$$\mathcal{E}(D; \boldsymbol{\theta}) = \sum_n -\log p(\mathbf{x}_n|\boldsymbol{\theta}) \quad (2)$$

and is minimize by the ML parameters. The Gaussian mixture model is extremely flexible and simply minimizing the above costfunction will lead to an “infinite overfit”. It is easily verified that the costfunction has a trivial (infinite) minimum attained by setting $\boldsymbol{\mu}_k = \mathbf{x}_k$ for $k = 1, \dots, K - 1$, and letting the corresponding covariances shrink to the zero matrix, while the remaining K ’th Gaussian is adapted to the ML fit of the remaining $N - K + 1$ datapoints. This solution is optimal for the training set, but unfortunately has a generalization error roughly equal to that of the single “background” Gaussian. To see this, let the

Figure E.76: Unpublished article (Nielsen and Hansen 1999). Page 2 of 7.

generalization error is defined as the limit

$$\Gamma(\theta) = \lim_{N \rightarrow \infty} \sum_{n=1}^N -\log p(x_n | \theta). \quad (3)$$

The ML mixture adapted on a finite dataset has a generalization error where the singular components do not contribute because the data points assigned to the singular datapoints in the training set together have zero measure. This instability has lead to much confusion in the literature and needs to be addressed carefully. Basically, there is no way to distinguish generalizable from non-generalizable solutions, if we only consider the likelihood function. The most common fix is to bias the component distributions so that they have a common covariance matrix, see e.g. [10]. Here we have decided to combine three approaches to ensure generalizability. First, we compute centers and covariances on different resamples of the data sets. Secondly, we make an exception rule for sparsely populated components — the covariance matrix defaults to the scaled full-sample covariance matrix. Finally we estimate the number of mixture components using the AIC criterion [1].

The algorithm which is a modified EM procedure [2], and is defined as follows for given K .

Algorithm: Generalizable Gaussian Mixture

Initialization

1. Compute the mean vector $\mu_0 = N^{-1} \sum_n x_n$.
2. Compute the covariance matrix of the data set: $\Sigma_0 = N^{-1} \sum_n (x_n - \mu_0)(x_n - \mu_0)^T$.
3. Initialize $\mu_k \sim \mathcal{N}(\mu_0, \Sigma_0)$.
4. Initialize $\Sigma_k = \Sigma_0$.
5. Initialize $P(k) = 1/K$.

Repeat until convergence

1. Compute $p(k|x_n)$ and assign x_n to the most likely component.
2. Split the data set in two equal parts D_μ, D_Σ .
3. For each k estimate μ_k on the points in D_μ assigned to component k .
4. For each k estimate Σ_k on the points in D_Σ assigned to component k . If the number of datapoints assigned to the k 'th component is less than the dimension, let $\Sigma_k = c \cdot \Sigma_0$, where c is determined so that the total variance of the component corresponds to the variance of the datapoints associated with it.
5. Estimate $P(k)$ as the frequency of assignments to component k .

3.1 Generalizable Gaussian mixture classifier

In pattern recognition we are interested in the joint density of patterns \mathbf{x} and class labels c , denoted by $p(\mathbf{x}, c)$

$$p(\mathbf{x}, c) = p(\mathbf{x}|c)P(c) \quad (4)$$

where $p(\mathbf{x}|c)$ is the class conditioned density and $P(c)$ is the marginal class probabilities. For a labeled dataset we design the classifier by adapting GGM's to each class separately. Hence, the joint density can be written

$$p(\mathbf{x}, c) = \sum_{k=1}^{K_c} p(\mathbf{x}|k)P(k|c)P(c). \quad (5)$$

Figure E.77: Unpublished article (Nielsen and Hansen 1999). Page 3 of 7.

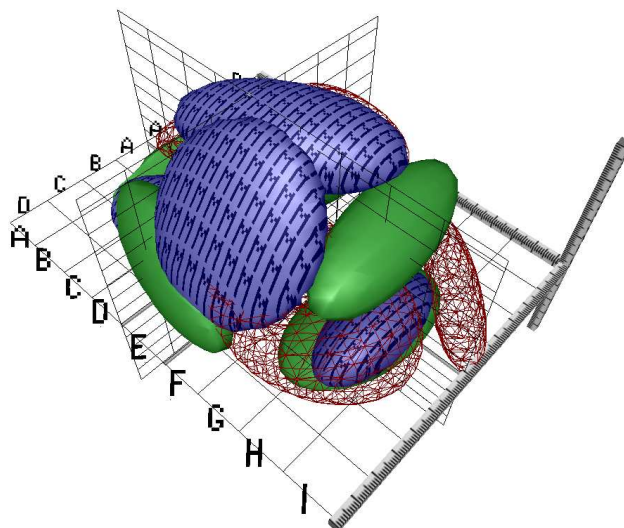


Figure 1: VRML screen shot (from the left side of the back of the brain) of the GGM model from the analysis of the effect of the behavioral domain. It shows isosurfaces in 3 class-conditional densities $p(\mathbf{x}|c)$. The wireframes correspond to the behavioral domain denoted perception in BrainMap, the surfaces to cognition apart from "M" textured surfaces being motion.

where $P(k|c)$ and K_c are the component frequencies and number components found for class c .

4 Results

We downloaded the entire paper and experiment webpages from the BrainMap homepage containing experiment variables and Talairach coordinates denoted locations. The front page of the web interface states that there are 225 papers, 771 experiments and 7683 locations. To each of the locations corresponds a modality, i.e., the type of scanner used to acquire the data. Furthermore, each location has one or more behavioral domains: Perception, cognition, motion, disease, drug and emotion. We used only the first behavioral domain in the list and confined us to the three classes: perception, cognition and motion. (For some of the experiments the modality and behavioral domain was not reported in the database and in the preliminary experiments reported here we simply excluded locations from these experiments.) While modeling the density of locations we found that some of the locations were strong outliers. Some of these foci were probably erroneous entries with decimal point errors. When excluding all locations from papers containing outliers or missing behavioral domain and modality we are left with approximately 3800

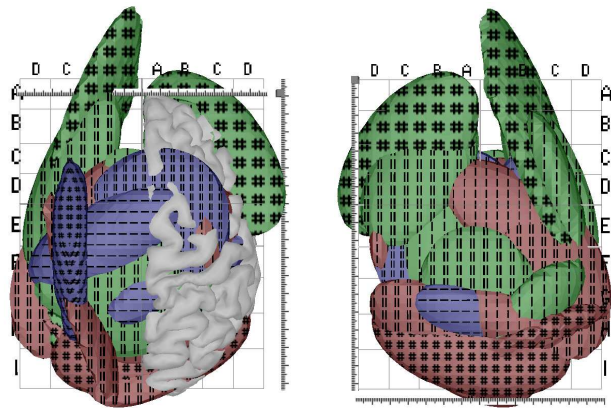


Figure 2: VRML screen shot from the top (left panel) and the bottom (right panel) of the brain. The frontal brain is up in the image. The red surfaces correspond to the behavioral domain denoted perception, the green to cognition and the blue being motion. Textured surfaces with # are derived from the density of fMRI locations, - are PET locations and = are PET-MRI locations. For visual guidance and reference we have included in the left panel also a surface reconstruction of the right cortex of the Visible Man [3].

Talairach coordinates.

We applied the GGM on the BrainMap data with the 3D Talairach coordinates as x and the modality or behavioral domain represented by label c . In figure 1 is shown the isosurface in the class-conditional densities $p(x|c)$ for $c \in \{\text{perception, cognition, motion}\}$. A characteristic in this view is the motion cluster in the left hemisphere (as e.g., compared to the right hemisphere) probably stemming from the popular use of the right hand in studies.

In a second experiment we combined behavioral domain and modality for the labels, i.e. with 9 classes in total. Figure 2 shows that the frontal part of the brain is dominated by the conjunction of cognitive and fMRI. We can only speculate about this: Differing spatial normalization procedures or fMRI motion artifacts could have affected the images or the fMRI scanner could be more sensitive in the frontal part. PET dominates in the inferior part of the brain, e.g., in the frontal regions due to susceptibility distortions.

The generalization in terms of label prediction is not high, which is due to the high overlap between classes:

	Test set rates	Base line probability	
Modality	0.17	0.16	
Behavior	0.47	0.52	(Figure 1)
Modality+Behavior	0.52	0.56	(Figure 2)

With the density model at hand we are able to pick a new functional neuroimaging study and automatically label the activation foci and give the amount of novelty in the experiment

Figure E.79: Unpublished article (Nielsen and Hansen 1999). Page 5 of 7.

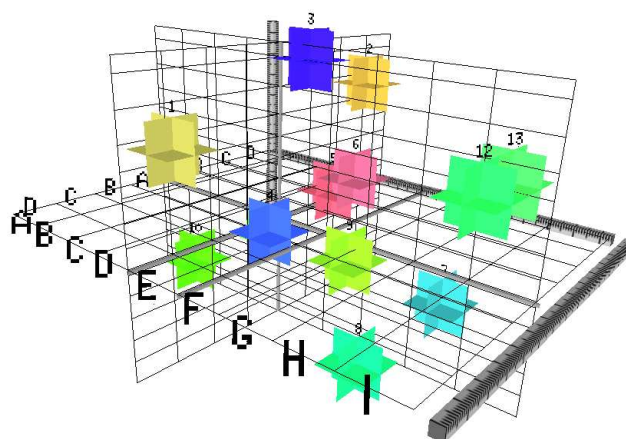


Figure 3: VRML Screen shot of a part of the results from [12]. The glyphs have been colored according to the posterior probability: The red component denote high probability for perception, green component for cognition, and blue component for motion.

(in fact, one location outlier was discovered in this way). In figure 3 is shown the results when the density model with labels from **behavioral domain** is applied on a saccadic eye-movement experiment [12]. The activation foci reported in the paper have automatically been labeled by the GGM model and the three components of the posterior probabilities have been piped to the red, green and blue color component. The highest posterior probability for a single **behavioral domain** is the third focus with the most probable label being motion. In the paper the foci has been denoted as the supplementary eye field, — an area associated with the control of eye movements.

5 Conclusion

Machine learning methods can assist the neuroscientists in quality control and provide context by summarizing large neuroimaging databases.

Acknowledgments

We thank Research Imaging Center, University of Texas Health Science Center at San Antonio for access to the BrainMap database. This paper has been supported by the Danish research councils through “THOR center for Neuroinformatics” and “Interdisciplinary Neuroresearch” and the American NIH “Human Brain Project” grant R01 DA09246 and P20 MH57180.

References

- [1] H. Akaike. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19:716–723, 1974.
- [2] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, 39:1–38, 1977.
- [3] Heather A. Drury and David C. Van Essen. *CARET User's Guide*. Washington University School of Medicine, St. Louis, Missouri, USA, September 1997.
- [4] Brian S. Everitt and Edward T. Bullmore. Mixture model mapping of brain activation in functional magnetic resonance images. *Human Brain Mapping*, 7(1):1–14, 1999.
- [5] Martha J. Farah and Geoffrey K. Aguirre. Imaging visual recognition: PET and fMRI studies of the functional anatomy of human visual recognition. *Trends in Cognitive Sciences*, 3(5):179–186, May 1999.
- [6] Peter T. Fox and Jack L. Lancaster. Neuroscience on the net. *Science*, 266(5187):994–996, November 1994.
- [7] Peter T. Fox, Jack L. Lancaster, Lawrence M. Parson, Jin-Hu Xiong, and Frank Zarnarripa. Functional volumes modeling: Theory and preliminary assessment. *Human Brain Mapping*, 5(4):306–311, 1997.
- [8] Peter T. Fox, Jack L. Lancaster, Lawrence M. Parsons, and Jin-Hu Xiong. Functional volumes modeling: Metanalytic models for statistical parametric imaging. In Lars Friberg, Albert Gjedde, Søren Holm, Niels A. Lassen, and Markus Nowak, editors, *Third International Conference on Functional Mapping of the Human Brain, NeuroImage*, volume 5, page S397. Academic Press, May 1997.
- [9] Peter T. Fox, Lawrence M. Parsons, and Jack L. Lancaster. Beyond the single study: functional/location metanalysis in cognitive neuroimaging. *Current Opinion in Neurobiology*, 8(2):178–187, April 1998.
- [10] Trevor Hastie and Robert Tibshirani. Discriminant analysis by Gaussian mixtures. *Journal of the Royal Statistical Society, Series B, Methodology*, 58(1):155–176, 1996.
- [11] Jack L. Lancaster, Peter T. Fox, Gwendolyn Davis, and Shawn Mikiten. BrainMap: A database of human functional brain mapping. In *The Fifth International Conference: Peace through Mind/Brain Science*, Hammamatsu, Japan, February 1994.
- [12] Ian Law, Claus Svarer, Egill Rostrup, and Olaf B. Paulson. Parieto-occipital cortex activation during self-generated eye movements in the dark. *Brain*, 121(11):2189–2200, November 1998.
- [13] M. Lepage, R. Habib, and E. Tulving. Hippocampal PET activations of memory encoding and retrieval: The HIPER model. *Hippocampus*, 8(4):313–322, 1998.
- [14] Michael I. Posner and Marcus E. Raichle. *Images of Mind*. Scientific American Library, 1997.
- [15] Brian D. Ripley. *Pattern Recognition and Neural Networks*. Cambridge University Press, 1996.
- [16] Gordon M. Shepherd, Jason S. Mirsky, Matthew D. Healy, Michael S. Singer, Emmanouil Skoufos, Michael S. Hines, Parkash M. Nadkarni, and Perry L. Miller. The Human Brain Project: neuroinformatics tools for integrating, searching and modeling multidisciplinary neuroscience data. *Trends in Neurosciences*, 21(11):460–468, November 1998.
- [17] J. Talairach and P. Tournoux. *Co-planar Stereotaxic Atlas of the Human Brain*. Thieme Medical Publisher Inc, New York, 1988.

Figure E.81: Unpublished article (Nielsen and Hansen 1999). Page 7 of 7

Functional Volumes Modeling using Kernel Density Estimation

F. Å. Nielsen and L. K. Hansen

Dept. of Mathematical Modelling, Technical University of Denmark, Lyngby, Denmark

Introduction

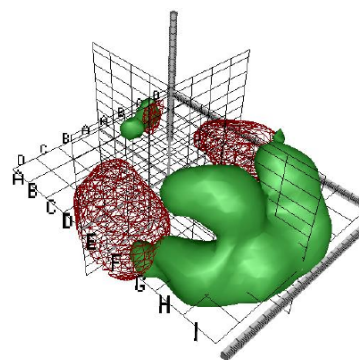
We describe a method based on kernel density estimation (also called Parzen window density estimation) for metaanalysis of brain maps, so-called functional volumes modeling (FVM) [1, 2]. We view FVM as the task of determining the conditional probability of activation foci given the design: paradigm, modality, number of subjects, etc. In [1, 2] the distribution of activation in connection with the mouth is modeled with a single Gaussian. The kernel density estimation (KDE) technique allows us to have a more flexible shape of the distribution, e.g., bimodal.

Method

KDE for FVM takes a simple multivariate distribution (the kernel) and places a local instance of it in each Talairach point. The resulting density is the sum of the local densities. We use a Gaussian with homogeneous variance as the kernel, optimizing the kernel width by leave-one-out (LOO) cross-validation. This procedure can be applied on any set of Talairach points, e.g.: The entire data from the BrainMap database [3] was downloaded from the Internet. From this data we extracted the foci and labeled those for which the field "Behavioral Domain" had only one value and were either "Audition" or "Vision", leaving 426 (60+366) 3D Talairach points. We constructed the probability density function conditioned on the label. It would be expected that the distribution of such a broad category would neither be very focused nor Gaussian. We also modeled the M1-mouth activation data assembled in [2]. The number of subjects in each experiment was not modeled.

Result

The optimization of the single parameter in the model — the kernel width — can be done fast using a Newton method. The figure shows the isosurface in the two probability density volumes rendered in 3D using VRML [4], the wireframe being for "Audition" and polygon for "Vision". The densities are not focused: With the threshold for the isosurface in the figure the probability mass within the isosurface was $P_{\text{audition}} = 0.17$ and $P_{\text{vision}} = 0.39$. The points with the highest densities for "Audition" are in the superior temporal lobe (the left area: -53,-19,4). The "Vision"-density shows a marked non-Gaussianity. Modeling the data from the foci mentioned in [2] reveals two distinct modes in the left and right M1-mouth area. It is not necessary to split the left and right set of foci before applying KDE.



Discussion

Contrary to [1, 2] we do not require any assumption of Gaussianity. On the other hand we are not able to model intersubject variability and the resulting distribution is dependent on the distribution of the number of subjects in each experiment. The ease of use of this method makes it a candidate for the set of standard techniques in FVM, where it can be used to form regions for explicit hypothesis testing in voxel-based studies. Other techniques might also be interesting in FVM: finite and infinite [5] Gaussian mixture models.

Acknowledgement

The Danish Research Councils and the Human Brain Project P20 MH57180.

References

1. Fox, P. T., et al., Human Brain Mapping, 1997, 5(4):306–311.
2. Fox, P. T., et al., Human Brain Mapping, 1999, 8(2–3):143–150.
3. Fox, P. T., Lancaster, J. L., Science, November 1994, 266(5187):994–996.
4. Nielsen, F. Å., Hansen, L. K., NeuroImage, vol. 7, 1998, S782.
5. Rasmussen, C. E., Advances in Neural Information Processing Systems 12, 2000.

Figure E.82: Unpublished abstract (Nielsen and Hansen 2000b).

Bibliography

- Abbott, D. and G. Jackson (2001, June). iBrain — software for analysis and visualisation of function MR images. *NeuroImage* **13**(6, part 2): S59. ISSN 1053-8119. <http://www.apnet.com/www/journal/hbm2001/10839.html>.
- Aguirre, G. K., E. Zarahn, and M. D'Esposito (1998a, October). The inferential impact of global signal covariates in functional neuroimaging analyses. *NeuroImage* **8**(3): 302–306. PMID: 9758743. <http://www.idealibrary.com/links/doi/10.1006/nimg.1998.0367>.
- Aguirre, G. K., E. Zarahn, and M. D'Esposito (1998b, November). The variability of human BOLD hemodynamic responses. *NeuroImage* **8**(4): 360–369. <http://www.idealibrary.com/links/artid/nimg.1998.0369>.
- Aho, A. V., R. Sethi, and J. D. Ullman (1986). *Compilers. Principles, Techniques, and Tools* (World Student Series Edition ed.). Addison-Wesley series in Computer Science. Addison-Wesley Publishing Company. ISBN 0201101947. “The Dragon Book”.
- Akaho, S., S. Hayamizu, O. Hasegawa, T. Yoshimura, and H. Asoh (1997, April). Multiple attribute learning with canonical correlation analysis and EM algorithm. Technical Report 97-8, Electrotechnical Laboratory, Ibaraki, Japan. <http://www.etl.go.jp/~akaho/papers/ETL-TR-97-8E.ps.gz>.
- Akaike, H. (1969). Fitting autoregressive models for prediction. *Annals of the Institute of Statistical Mathematics* **21**: 243–247.
- Akaike, H. (1970). Statistical predictor identification. *Annals of the Institute of Statistical Mathematics* **22**: 203–217.
- Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov and F. Cásiki (Eds.), *Second International Symposium on Information Theory*, Budapest, pp. 267–281. Akademiai Kiadó. Reprinted in *Breakthroughs in Statistics*.
- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control* **19**: 716–723.
- Akaike, H. (1977). On entropy maximisation principle. In P. R. Krishnaiah (Ed.), *Applications in Statistics*, pp. 27–41. Amsterdam: North-Holland.
- Akaike, H. (1978). A Bayesian analysis of the minimum AIC procedure. *Annals of the Institute of Statistical Mathematics* **30A**: 9–14.
- Akeley, K. (1993). RealityEngine graphics. In *Computer Graphics SIGGRAPH'93*, pp. 109–116.
- Allison, T., G. McCarthy, A. Nobre, A. Puce, and A. Belger (1994, September-October). Human extrastriate visual cortex and the perception of faces, words, numbers, and colors. *Cerebral Cortex* **4**(5): 544–554. PMID: 7833655.
- Allison, T., A. Puce, and G. McCarthy (2000). Social perception from visual cues: role of the STS region. *Trends in Cognitive Sciences* **4**(7): 267–278.
- Almeida, R. and A. Ledberg (2001). A spatially constrained clustering algorithm with no prior knowledge of the number of clusters. pp. S61. <http://www.apnet.com/www/journal/hbm2001/10074.html>.
- Almind, T. C. and P. Ingwersen (1997, September). Informetric analyses on the world wide web: Methodological approaches to “webometrics”. *Journal of Documentation* **53**(4): 404–426.
- Altschul, S. F., W. Gish, E. W. Myers, and D. J. Lipman (1990, October). Basic local alignment search tool. *Journal of Molecular Biology* **215**(3): 403–410. PMID: 2231712.
- Amari, S. (1998). Natural gradient works efficiently in learning. *Neural Computation* **10**(2): 251–276.

- Amari, S. (1999, November). Natural gradient learning for over- and under-complete bases in ICA. *Neural Computation* **11**(8): 1875–1883. PMID: 10578035.
- Amari, S. (2000, September). Estimating functions of independent component analysis for temporally correlated signals. *Neural Computation* **12**(9): 2083–2107. PMID: 10976140.
- Amunts, K., A. Malikovic, H. Mohlberg, T. Schormann, and K. Zilles (2000, January). Brodmann's areas 17 and 18 brought into stereotaxic space-where and how variable? *NeuroImage* **11**(1): 66–84. PMID: 10686118.
- Amunts, K., A. Schleicher, U. Bürgel, H. Mohlberg, H. B. Uylings, and K. Zilles (1999, September). Broca's region revisited: cytoarchitecture and intersubject variability. *Journal of Comparative Neurology* **412**(2): 319–341. PMID: 10441759.
- Andersen, A. H., D. M. Gash, and M. J. Avison (1999, July). Principal component analysis of the dynamic response measured by fMRI: A generalized linear systems framework. *Magnetic Resonance Imaging* **17**(6): 795–815. PMID: 10402587.
- Andersen, L. N., J. Larsen, L. K. Hansen, and M. Hintz-Madsen (1997, September). Adaptive regularization of neural classifiers. In J. Principe (Ed.), *Proceedings of the IEEE Workshop on Neural Networks for Signal Processing VII, Florida, USA*, Piscataway, New Jersey, pp. 24–33. IEEE. <ftp://eivind.imm.dtu.dk/dist/1997/nonboe.nnsp97.ps.Z>.
- Anderson, T. W. (1984). *An Introduction to Multivariate Statistical Analysis* (Second ed.). Wiley series in probability and Mathematical Statistics. New York: John Wiley & Sons. ISBN 0-471-88987-3.
- Andersson, J. L. R. (1997, November). How to estimate global activity independent of changes in local activity. *NeuroImage* **6**(4): 237–244. PMID: 9453855. <http://www.idealibrary.com/links/citation/1053-8119/6/237>.
- Andersson, J. L. R. (1999a, April). Re: Global values. SPM mailing list. <http://www.jiscmail.ac.uk/cgi-bin/wa.exe?A2=ind9904&L=spm&F=&S=&P=532>.
- Andersson, J. L. R. (1999b, March). Re: Normalization. SPM mailing list. <http://www.jiscmail.ac.uk/cgi-bin/wa.exe?A2=ind9903&L=spm&D=0&P=8070>.
- Andersson, J. L. R. (2000, December). The ghost in SPM. SPM mailing list. <http://www.mailbase.ac.uk/lists/spm/>.
- Andersson, J. L. R. (2001a, May). Modelling geometric deformations in EPI time series. *NeuroImage* **13**(5): 903–919. PMID: 11304086. <http://www.fil.ion.ucl.ac.uk/spm/papers/Unwarp/Unwarp.pdf>.
- Andersson, J. L. R. (2001b, April). New toolbox. SPM Mailing list. <http://www.jiscmail.ac.uk/cgi-bin/wa.exe?A2=ind0104&L=spm&F=&S=&P=26856>.
- Andersson, J. L. R., J. Ashburner, and K. J. Friston (2001, June). A global estimator unbiased by local changes. *NeuroImage* **13**(6): 1193–1205. PMID: 11352625. <http://www.idealibrary.com/links/citation/1053-8119/13/1193>.
- Andersson, J. L. R., A. Sundin, and S. Valind (1995, July). A method for coregistration of PET and MR brain images. *Journal of Nuclear Medicine* **36**(7): 1307–1315. PMID: 7790961.
- Ardekani, B., M. Braun, B. F. Hutton, I. Kanno, and H. Iida (1995, July-August). A fully automatic multimodality image registration algorithm. *Journal of Computer Assisted Tomography* **19**(4): 615–623. PMID: 7622696.
- Ardekani, B. A., J. Kershaw, K. Kashikura, and I. Kanno (1999, February). Activation detection in functional MRI using subspace modeling and maximum likelihood estimation. *IEEE Transaction on Medical Imaging* **18**(2): 101–114. PMID: 10232667.
- Armitage, P. and G. Berry (1994). *Statistical Methods in Medical Research* (Third ed.). Blackwell Science. ISBN 0-632-03695.
- Armstrong, J. S. (1982, June). Barriers to scientific contributions: The author's formula. *Behavioral and Brain Sciences* **5**: 197–199. <http://fourps.wharton.upenn.edu/ideas/pdf/barriers.pdf>.
- Armstrong, J. S. and R. Hubbard (1991). Does the need for agreement among reviewers inhibit the publication of controversial findings? *Behavioral and Brain Sciences* **14**: 136–137. <http://fourps.wharton.upenn.edu/ideas/pdf/agree.pdf>.
- Arndt, S., T. Cizadlo, D. O'Leary, S. Hold, and N. C. Andreasen (1996, June). Normalizing counts and cerebral blood flow intensity in functional imaging studies of the human brain. *NeuroImage* **3**(3, Part 1 of 2): 175–184.
- Arnfred, S., A. C. N. Chen, D. Eder, B. Glenthøj, and R. Hemmingsen (2000). Proprioceptive evoked potentials in man: Cerebral responses to changing weight loads on the hand. *Neuroscience Letters* **288**: 111–114.

- Arnold, J. B., J. S. Liow, K. A. Schaper, J. J. Stern, J. G. Sled, D. W. Shattuck, A. J. Worth, M. S. Cohen, R. M. Leahy, J. C. Mazziotta, and D. A. Rottenberg (2001, May). Qualitative and quantitative evaluation of six algorithms for correcting intensity nonuniformity effects. *NeuroImage* **13**(5): 931–943. PMID: 11304088.
- Ashburner, J. and K. J. Friston (1996, June). Fully three-dimensional nonlinear spatial normalization. See Belliveau, Fox, Kennedy, Rosen, and Ungeleider (1996), pp. S111. ISSN 1053–8119.
- Ashburner, J. and K. J. Friston (1997). Multimodal image coregistration and partitioning — a unified framework. *NeuroImage* **6**(3): 209–217. PMID: 9344825.
- Ashburner, J. and K. J. Friston (1999). Nonlinear spatial normalization using basis functions. *Human Brain Mapping* **7**(4): 254–266.
- Ashburner, J., J. Haslam, C. Taylor, V. J. Cunningham, and T. Jones (1996). A cluster analysis approach for the characterization of dynamic PET data. Chapter 59, pp. 301–306. San Diego: Academic Press.
- Ashburner, J., P. Neelin, D. L. Collins, A. C. Evans, and K. J. Friston (1997, November). Incorporating prior knowledge into image registration. *NeuroImage* **6**(4): 344–352. ISSN 1053-8119. PMID: 9417976. <http://www.idealibrary.com/links/citation/1053-8119/6/344>.
- Asoh, H. and O. Takechi (1994). An approximation of nonlinear canonical correlation analysis by multilayer perceptrons. In *Proceedings of ICANN'94*, pp. 713–716.
- Aston, J. A. D., K. J. Worsley, and R. N. Gunn (2001, June). RPM STATISTICS — a statistical tool for receptor parametric mapping. *NeuroImage* **13**(6, part 2): S65. <http://www.apnet.com/www/journal/hbm2001/9786.html>.
- Atkinson, R. L., R. C. Atkinson, E. E. Smith, D. J. Bem, and E. R. Hilgard (1990). *Introduction to Psychology* (Tenth ed.). San Diego: Harcourt Brace Jovanovich. ISBN 0-15-543688-0.
- Attias, H. and C. E. Schreiner (1998). Blind source separation and deconvolution: The dynamic component analysis algorithm. *Neural Computation* **10**: 1373–1424.
- Baldi, P. and K. Hornik (1989). Neural networks for principal component analysis: Learning from examples without local minima. *Neural Networks* **2**: 53–58.
- Balslev, D., I. Law, S. A. Frutiger, J. J. Sidtis, F. Å. Nielsen, T. B. Christiansen, S. C. Strother, C. Svarer, D. A. Rottenberg, O. B. Paulson, and L. K. Hansen (2000, May). Visuomotor skill learning: a PET study of mirror tracing using cluster analysis and statistical parametric mapping. See Fox and Lancaster (2000), pp. S371. ISSN 1053-8119.
- Balslev, D., F. Å. Nielsen, S. A. Frutiger, J. J. Sidtis, T. B. Christiansen, C. Svarer, S. C. Strother, D. A. Rottenberg, L. K. Hansen, O. B. Paulson, and I. Law (2001). Cluster analysis of activity-time series in motor learning. Under review.
- Baluja, S. and D. A. Pomerleau (1995). Using a saliency map for active spatial selective attention: Implementation & initial results. See Tesauro, Touretzky, and Leen (1995), pp. 451–458.
- Bandettini, P. A., A. Jesmanowicz, E. C. Wong, and J. S. Hyde (1993, August). Processing strategies for time-course data sets in functional MRI of the human brain. *Magnetic Resonance in Medicine* **20**(2): 615–173. PMID: 8366797.
- Banfield, J. D. (1993, September). Model-based gaussian and non-gaussian clustering. *Biometrics* **49**: 803+.
- Barinaga, M. (1997, April). What makes brain neurons run. *Science* **276**(5310): 196–198. PMID: 9132940.
- Barker, A. T., R. Jalinous, and I. L. Freeston (1985, May). Non-invasive magnetic stimulation of human motor cortex. *Lancet* **1**(8437): 1106–1107. PMID: 2860322.
- Barlow, H. B. (1989). Unsupervised learning. *Neural Computation* **1**: 295–311.
- Baumgartner, R., L. Ryner, R. Summers, M. Jarmasz, and R. L. Somorjai (1999, June). Exploratory data analysis results in fMRI may be corroborated by statistical inferential methods. See Rosen, Seitz, and Volkman (1999), pp. S47. ISSN 1053–8119.
- Baune, A., F. T. Sommer, M. Erb, D. Wildgruber, B. Kardatzki, G. Palm, and W. Grodd (1999, May). Dynamical cluster analysis of cortical fMRI activation. *NeuroImage* **9**(5): 477–489. PMID: 10329287.
- Bayer, A. E., J. C. Smart, and G. W. McLaughlin (1990). Mapping intellectual structure of a scientific subfield through author cocitations. *Journal of the American Society for Information Science* **41**(6): 433–443.

- Becker, J. T., M. A. Mintun, D. J. Diehl, J. Dobkin, A. Martidis, D. C. Madoff, and S. T. Dekosky (1994). Functional neuroanatomy of verbal free recall: A replication study. *Human Brain Mapping* **1**(4): 284–292.
- Becker, S. (1992). *An Information-theoretic Unsupervised Learning Algorithm for Neural Networks*. Ph. D. thesis, Graduate Department of Computer Science, University of Toronto. <http://www.science.mcmaster.ca/Psychology/becker/Publications.html>.
- Becker, S. and G. E. Hinton (1992, January). A self-organizing neural network that discovers surfaces in random-dot stereograms. *Nature* **355**(6356): 161–163.
- Becker, S. and Y. Le Cun (1988). Improving the convergence of back-probagation learning with second order methods. See Tourtzky, Hinton, and Sejnowski (1988), pp. 29–37.
- Beckman, R. J. and R. D. Cook (1983, May). Outlier.....s. *Technometrics* **25**(2): 119–149. Discussions 150–161.
- Beckmann, C. F., J. A. Noble, and S. M. Smith (2000, May). Artefact detection in FMRI data using independent component analysis. See Fox and Lancaster (2000), pp. S614. ISSN 1053-8119.
- Bell, A. J. and T. J. Sejnowski (1995). An information maximisation approach to blind separation and blind deconvolution. *Neural Computation* **7**(6): 1129–1159. <ftp://ftp.cnl.salk.edu/pub/tony/bell.blind.ps.Z>. ResearchIndex: <http://citeseer.nj.nec.com/bell95informationmaximization.html>.
- Bell, G., A. Parisi, and M. Pesce (1995, November). The virtual reality modeling language. Internet: <http://www.vrml.org>. VRML 1.0 Specification.
- Bellgowan, P. S. F., Z. Saad, and P. A. Bandettini (2001, June). Voxel-wise estimation of hemodynamic onset delays during a lexical decision task. *NeuroImage* **13**(6, part 2): S503. <http://www.apnet.com/www/journal/hbm2001/11775.html>.
- Belliveau, J., P. Fox, D. Kennedy, B. Rosen, and L. Ungerleider (Eds.) (1996, June). *Second International Conference on Functional Mapping of the Human Brain*, Volume 3. Academic Press. ISSN 1053–8119.
- Belliveau, J. W., D. N. Kennedy Jr, R. C. McKinstry, B. R. Buchbinder, R. M. Weisskoff, M. S. Cohen, J. M. Vevea, T. J. Brady, and B. R. Rosen (1991, November). Functional mapping of the human visual cortex by magnetic resonance imaging. *Science* **254**(5032): 716–719.
- Beltrami, E. (1873). Sulle funzioni bilineari. *Gionale di Matematiche* **11**: 98–106.
- Ben-Israel, A. and T. N. E. Greville (1980). *Generalized Inverses: Theory and Applications* (Reprint edition ed.). Huntington, New York: Robert E. Krieger Publishing Company.
- Bengio, Y. (2000, August). Gradient-based optimization of hyper-parameters. *Neural Computation* **12**(8): 1889–1900.
- Benjamini, Y. and Y. Hochberg (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B* **57**: 289–300.
- Berger, H. (1929). Über das elektrenkephalogramm des menschen. *Archiv für Psychiatrie und Nervenkrankheiten* **87**: 527–570.
- Berger, J. O., B. Liseo, and R. L. Wolpert (1999). Integrated likelihood methods for eliminating nuisance parameters (with discussion). *Statistical Science* **14**: 1–28. <http://ftp.isds.duke.edu/WorkingPapers/97-01.ps>. ResearchIndex: <http://citeseer.nj.nec.com/berger99integrated.html>.
- Berks, G., G. Pohl, and D. G. v. Keyserlingk (2001). 3D-VIEWER: An atlas-based system for individual and statistical investigations of the human brain. *Methods of Information in Medicine* **40**: 170–171.
- Berners-Lee, T., R. Cailliau, J.-F. Groff, and B. Pollermann (1992, Spring). World-Wide Web: The information universe. *Electronic Networking: Research, Applications and Policy* **2**(1): 52–58. http://www.w3.org/History/1992/ENRAP/Article_9202.ps. ResearchIndex: <http://citeseer.nj.nec.com/berners-lee92worldwide.html>.
- Berners-Lee, T., R. Cailliau, A. Luotonen, H. Nielsen, and A. Secret (1994). The world-wide web. *Communications of the ACM* **37**(8): 76–82.
- Bezdek, J. C. (1981). *Pattern Recognition with Fuzzy Objective Function Algorithms*. New York: Plenum Press.
- Bhansali, B. J. and D. Y. Downham (1977). Some properties of the order of an autoregressive model selected by a generalization of Akaike's FPE criterion. *Biometrika* **54**: 547–551.

- Bichteler, J. and E. A. Eaton III (1980, July). The combined use of bibliometric coupling and cocitation for document retrieval. *Journal of the American Society for Information Science* **31**(4).
- Biernacki, C., G. Celeux, and G. Govaert (2000). Assessing a mixture model for clustering with the integrated completed likelihood. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **22**: 719–725. ISSN 01628828.
- Bilke, S. and C. Peterson (2001). Topological properties of citation and metabolic networks. Technical Report LU TP 01-09, Department of Theoretical Physics, Lund University, Lund, Sweden. http://www.thep.lu.se/pub/Preprints/01/lu_tp_01_09.ps.gz. To appear in Physical Review E.
- Binder, J. R., J. A. Frost, T. A. Hammeke, P. S. Bellgowan, S. M. Rao, and R. W. Cox (1999, January). Conceptual processing during the conscious resting state. a functional mri study. *Journal of Cognitive Neuroscience* **11**(1): 80–95. PMID: 9950716.
- Binder, J. R., S. M. Rao, T. A. Hammeke, J. A. Frost, P. A. Bandettini, and J. S. Hyde (1994, July). Effects of stimulus rate on signal response during functional magnetic resonance imaging of auditory cortex. *Brain Research. Cognitive Brain Research* **2**(1): 31–38. PMID: 7812176.
- Birn, R. M. and P. A. Bandettini (2001, June). Estimated BOLD impulse response depends on stimulus on/off ratio. *NeuroImage* **13**(6): S971. <http://www.apnet.com/www/journal/hbm2001/11342.html>.
- Birrell, A. and P. MacJones (2000, February). pstotext. Internet. <http://www.research.compaq.com/SRC/virtualpaper/pstotext.html>.
- Bishop, C. M. (1994, August). Novelty detection and neural network validation. *IEE Proceedings — Vision, Image and Signal Processing* **141**(4): 217–222. ResearchIndex: <http://citeseer.nj.nec.com/bishop94novelty.html>. Document No. 19941330.
- Bishop, C. M. (1995a). *Neural Networks for Pattern Recognition*. Oxford University Press. ISBN 0-19-853864-2.
- Bishop, C. M. (1995b). Training with noise is equivalent to Tikhonov regularization. *Neural Computation* **7**: 108–116.
- Bishop, C. M. (1999). Bayesian PCA. See Kearns, Solla, and Cohen (1999), pp. 382–388. ISBN 0-262-11245-0. NIPS-11.
- Biswal, B. B. and J. L. Ulmer (1999, March/April). Blind source separation of multiple signal sources of fMRI data sets using independent component analysis. *Journal of Computer Assisted Tomography* **23**(2): 265–271. PMID: 10096335.
- Björnsson, C. H. (1971). *Læsbarhed*. Copenhagen: GEC.
- Bloch, F., W. W. Hansen, and M. Packard (1946). Nuclear induction. *Physical Review* **69**: 127.
- Bly, B. M., D. Rebbecki, G. Grasso, and S. J. Hanson (2001, June). A peer-to-peer database for brain imaging data. *NeuroImage* **13**(6): S82. <http://www.apnet.com/www/journal/hbm2001/11785.html>.
- Bodurka, J. and P. A. Bandettini (2001, June). Toward direct mapping of neuronal activity: MRI detection of ultra weak transient magnetic field changes. *NeuroImage* **13**(6, part 2): S964. <http://www.apnet.com/www/journal/hbm2001/10370.html>.
- Boissonnat, J.-D. and B. Geiger (1993). Three dimensional reconstruction of complex shapes based on the Delaunay triangulation. In R. S. Acharya and D. B. Goldgof (Eds.), *Biomedical Image Processing and Biomedical Visualization*, Volume 1905, pp. 964–975. SPIE.
- Bollacker, K., S. Lawrence, and C. L. Giles (1998). CiteSeer: An autonomous web agent for automatic retrieval and identification of interesting publications. In *Agents '98, 2nd International ACM Conference on Autonomous Agents*, pp. 116. ACM.
- Borga, M. (1998, May). *Learning Multidimensional Signal Processing*. Ph. D. thesis, Linköping University, Linköping, Sweden. ftp://ftp.imt.liu.se/pub/bilddb/Theses/PhDTheses/M_Borga_thesis.ps.gz.
- Born, A. P. (1998). *The function of the Primary Visual Cortex During Early Human development: An MRI Study*. Ph. D. thesis, Faculty of Medicine, University of Copenhagen.
- Born, A. P., I. Law, G. Wildschjötz, O. B. Paulson, and H. C. Lou (2001, July). Paradoxical rCBF decreases during visual stimulation in stage 3/4 sleep detected by H₂ ¹⁵O-PET. *NeuroImage* **13**(6): S1132. <http://www.apnet.com/www/journal/hbm2001/9939.html>.
- Born, A. P., T. E. Lund, L. G. Hanson, P. L. Steensgaard, P. L. Madsen, E. Rostrup, and H. C. Lou (2000a, April). Negative visually induced BOLD response during human adult slow-wave sleep. In *Proceedings of ISMRM*, pp. 498. International society for Magnetic Resonance in Imaging.

- Born, A. P., M. J. Miranda, E. Rostrup, P. B. Toft, B. Peitersen, H. B. W. Larsson, and H. C. Lou (2000b, February). Functional magnetic resonance imaging of the normal and abnormal visual system in early life. *Neuropediatrics* **31**(1): 24–32. PMID: 10774992.
- Born, A. P., E. Rostrup, H. Leth, B. Peitersen, and H. C. Lou (1996, February). Change of visually induced cortical activation patterns during development. *Lancet* **347**(9000): 543. PMID: 8596290.
- Borodin, A., G. O. Roberts, J. S. Rosenthal, and P. Tsaparas (2001). Finding authorities and hubs from link structures on the world wide web. In *Tenth International World Wide Web Conference*. ResearchIndex: <http://citeseer.nj.nec.com/borodin00finding.html>.
- Botafofo, R. A. and B. Scheiderman (1991, December). Identifying aggregates in hypertext structures. In *Proceedings of Hypertext '91*, pp. 63–74. <ftp://ftp.cs.umd.edu/pub/papers/papers/ncstrl.umcp/CS-TR-2650/CS-TR-2650.ps.Z>.
- Bottou, L. and Y. Bengio (1995). Convergence properties of the K-means algorithm. See Tesauro, Touretzky, and Leen (1995), pp. 585–592.
- Boulanouar, K., J. F. Demonet, I. Berry, F. Chollet, C. Manelfe, and P. Celsis (1996, June). fMRI study of the motor system dynamics using the deconvolved evoked response of activated pixels. See Belliveau, Fox, Kennedy, Rosen, and Ungeleider (1996), pp. S376. ISSN 1053–8119.
- Bowden, D. M. and R. F. Martin (1995). NeuroNames brain hierarchy. *NeuroImage* **2**(1): 63–84. ISSN 1053-8119. PMID: 9410576.
- Box, G. E. P. (1954). Some theorems on quadratic forms applied in the study of analysis of variance problems, ii. effects of inequality of variance and of correlation between errors in the two-way classification. *The Annals of Mathematical Statistics* **25**: 484–498.
- Box, G. E. P. and G. C. Tiao (1992). *Bayesian Inference in Statistical Analysis* (Wiley classics library edition ed.). John Wiley & Sons, Inc.
- Boynton, G. M., S. A. Engel, G. H. Glover, and D. J. Heeger (1996, July). Linear systems analysis of functional magnetic resonance imaging in human V1. *The Journal of Neuroscience* **16**(13): 4207–4221.
- Brahe, T. (1602). *Astronomiae instauratae progymnasmata*. Uraniborg, Denmark.
- Brent, R. P. (1973). *Algorithms for Minimization without derivatives*. Englewood Cliffs, New Jersey: Prentice-Hall.
- Brett, M. (1999, August). The MNI brain and the Talairach atlas. <http://www.mrc-cbu.cam.ac.uk/Imaging/-mnispace.html>.
- Brett, M. (2000, August). Slice display. Internet. http://www.mrc-cbu.cam.ac.uk/Imaging/display_slices.html.
- Brett, M., K. Christoff, R. Cusack, and J. L. Lancaster (2001, June). Using the Talairach atlas with the MNI template. *NeuroImage* **13**(6): S85. <http://www.apnet.com/www/journal/hbm2001/10661.html>.
- Bretthorst, G. L. (1992, July). Bayesian interpolation and deconvolution. Technical Report CR-RD-AS-92-4, Department of Chemistry, Washington University, St. Louis, Missouri. <http://bayes.wustl.edu/glb/deconvolution.ps.gz>. ResearchIndex: <http://citeseer.nj.nec.com/237656.html>.
- Bridle, J. S. (1990). Probabilistic interpretation of feedforward classification network outputs, with relationships to statistical pattern recognition. In F. Fogelman Soulié and J. Héroult (Eds.), *Neurocomputing: Algorithms, Architectures and Application*, pp. 227–236. New York: Springer-Verlag.
- Brin, S. and L. Page (1998). Anatomy of a large-scale hypertextual web search engine. In *Proc. 7th International World Wide Web Conference*.
- Bro, R. (1998). *Multi-way Analysis in the Food Industry. Models, Algorithms and Applications*. Ph. D. thesis, University of Amsterdam, The Netherlands. <http://www.models.kvl.dk/users/rasmus/thesis/thesis.html>.
- Bro-Nielsen, M. (1996, August). *Medical Image Registration and Surgery Simulation*. Ph. D. thesis, Department of Mathematical Modelling, Technical University of Denmark, Lyngby, Denmark. http://www.anamedic.com/people/bro/phd_dissertation.htm. IMM-PHD-1996-25.
- Broca, P. (1960). Remarks on the seat of the faculty of articulate language, followed by an observation of aphemia. In *Some Papers on the Cerebral Cortex*, pp. 49–72. Springfield, Illinois: Charles C. Thomas Publisher.

- Brockway, J. P. (2000, June-August). Two functional magnetic resonance imaging f(mri) tasks that may replace the gold standard, wada testing, for language lateralization while giving additional localization information. *Brain Cogn* **43**(1-3): 57–59.
- Broder, A., R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins, and J. Wiener (2000). Graph structure in the web. In *Proceedings of the Ninth International World Wide Web Conference*. <http://www9.org/w9cdrom/160/160.html>.
- Brodmann, K. (1909). *Vergleichende Lokalisationslehre der Grosshirnrinde in ihren Prinzipien dargestellt auf Grund des Zellenbaues*. Leipzig: Barth.
- Brooks, R. J. and M. Stone (1994). Joint continuum regression for multiple predictands. *Journal of the American Statistical Association* **89**: 1374–1377.
- Brown, M. C. (1999). Audition. Chapter 27, pp. 791–820. San Diego, California: Academic Press.
- Brown, P. J. (1977). Centering and scaling in ridge regression. *Technometrics* **19**: 35–36.
- Broyden, C. G. (1967). Quasi-Newton methods and their application to function minimization. *Maths. Comput.* **21**: 368–381.
- Broyden, C. G. (1970). The convergence of a class of double-rank minimization algorithms. parts i and ii. *Journal of the Institute of Mathematics and Its Applications* **6**: 76–90 and 222–231.
- Bryan, R. N., C. Davatzikos, and M. Vaillant (1995, June). Creation of population-based anatomic atlases with a brain image database (BRAID). In *First International Conference on Functional Mapping of the Human Brain, Paris*.
- Büchel, C., A. P. Holmes, G. Rees, and K. J. Friston (1998, August). Characterizing stimulus-response functions using nonlinear regressors in parametric fMRI experiments. *NeuroImage* **8**(2): 140–148. PMID: 9740757.
- Büchel, C., R. J. S. Wise, C. J. Mummary, J.-B. Poline, and K. J. Friston (1996, August). Nonlinear regression in parametric activation studies. *NeuroImage* **4**(1): 60–66. PMID: 9345497.
- Buckner, R. L. (1998). Event-related fMRI and the hemodynamic response. *Human Brain Mapping* **6**(5-6): 373–377. PMID: 9788075. <http://www3.interscience.wiley.com/cgi-bin/abstract/79015/START>.
- Buckner, R. L. and S. E. Petersen (1996). What does neuroimaging tell us about the role of prefrontal cortex in memory retrieval. *Seminars in the Neuroscience* **8**: 47–55.
- Buckner, R. L., M. E. Raichle, F. M. Miezin, and S. E. Petersen (1996, October). Functional anatomic studies of memory retrieval for auditory words and visual pictures. *The Journal of Neuroscience* **16**(19): 6219–6235.
- Bullmore, E., M. Brammer, S. C. R. Williams, S. Rabe-Hesketh, N. Janot, A. David, J. Mellers, R. Howard, and P. Sham (1996a, February). Statistical methods of estimation and inference for functional MR image analysis. *Magnetic Resonance in Medicine* **35**(2): 261–277. PMID: 8622592.
- Bullmore, E. T., S. Rabe-Hesketh, R. G. Morris, S. C. Williams, L. Gregory, J. A. Gray, and M. J. Brammer (1996b, August). Functional magnetic resonance image analysis of a large-scale neurocognitive network. *NeuroImage* **4**(1): 16–33. PMID: 9345494. <http://www.idealibrary.com/links/doi/10.1006/nimg.1996.0026>.
- Buntine, W. (1996). A guide to the literature on learning probabilistic networks from data. *IEEE Transaction on Knowledge and Data Engineering* **8**(2): 195–210. <http://www-cad.eecs.berkeley.edu/~wray/graphbib.ps.Z>. ResearchIndex: <http://citeseer.nj.nec.com/buntine96guide.html>.
- Burnham, A. J., J. F. MacGregor, and R. Viveros (1999). A statistical framework for multivariate latent variable regression methods based on maximum likelihood. *Journal of Chemometrics* **13**: 49–65.
- Burnham, A. J., R. Viveros, and J. F. MacGregor (1996). Frameworks for latent variable multivariate regression. *Journal of Chemometrics*: 31–45.
- Burns, G. A. P. C. (1998). Neuroscholar 1.00, a neuroinformatics databasing website. In J. M. Bower (Ed.), *Neurocomputing: Proceedings of the Computational Neuroscience Conference*. Elsevier. <http://neuroscholar.usc.edu/>.
- Burt, C. (1950). The factorial analysis of qualitative data. *British Journal of Statistical Psychology (Stat. Sec.)* **3**: 166–185.

- Busatto, G., R. J. Howard, Y. Ha., M. Brammer, I. Wright, P. W. R. Woodruff, A. Simmons, S. C. R. Williams, A. S. David, and E. T. Bullmore (1997, August). A functional magnetic resonance imaging study of episodic memory. *NeuroReport* **8**(12): 2671–2675. PMID: 9295098.
- Bush, G., J. A. Frazier, S. L. Rauch, L. J. Seidman, P. J. Whalen, M. A. Jenike, B. R. Rosen, and J. Biederman (1999, June). Anterior cingulate cortex dysfunction in attention deficit/hyperactivity disorder revealed by fMRI and the counting stroop. *Biol. Psychiatry* **45**(12): 1542–1552. PMID: 10376114.
- Bush, G., P. Luu, and M. I. Posner (2000, June). Cognitive and emotional influences in anterior cingulate cortex. *Trends in Cognitive Sciences* **4**(6): 215–222. PMID: 10827444.
- Bush, V. (1945, July). As we may think. *The Atlantic Monthly* **1**: 101–108. <http://www.theatlantic.com/unbound/flashbks/computer/bushf.htm>.
- Buxton, R. B. (2001, June). The elusive initial dip. *NeuroImage* **13**(6): 63–958. Commentary.
- Buxton, R. B. and L. R. Frank (1997, January). A model for the coupling between cerebral blood flow and oxygen metabolism during neuronal stimulation. *Journal of Cerebral Blood Flow and Metabolism* **17**(1): 64–72. PMID: 8978388.
- Buxton, R. B., E. C. Wong, and L. R. Frank (1998). Dynamics of blood flow and oxygenation changes during brain activation: The balloon model. *Magnetic Resonance in Medicine* **39**: 855–864.
- Cabeza, R. and L. Nyberg (2000). Imaging cognition II: An empirical review of 275 PET and fMRI studies. *Journal of Cognitive Neuroscience* **12**(1): 1–47.
- Cabral, B. K. and J. Foran (1999, December). System and method of performing tomographic reconstruction and volume rendering using texture mapping. United States Patent No. 6,002,738. <http://www.uspto.gov/patft/>, http://www.delphion.com/details?pn=US06002738__.
- Cao, J. and K. J. Worsley (1999). The detection of local shape changes via the geometry of Hotelling's t^2 fields. *Annals of Statistics* **27**(3): 925–942. <http://cm.bell-labs.com/cm/ms/departments/sia/cao/ps/hotel8.ps.gz>.
- Cao, J. and K. J. Worsley (2001). Applications of random fields in human brain mapping. In M. Moore (Ed.), *Spatial Statistics: Methodological Aspects and Applications*, Volume 159 of *Lecture notes in Statistics*, Chapter 8, pp. 170–182. New York: Springer. <http://www.math.mcgill.ca/~keith/ws/ws.abstract.html>.
- Card, S. K., J. D. Mackinlay, and B. Shneiderman (Eds.) (1999). *Readings in information visualization: Using vision to think*. San Francisco: Morgan Kaufmann Publishers.
- Carey, R., G. Bell, and C. Marrin (1997). ISO/IEC 14772-1:1997 Virtual Reality Modeling Language (VRML97). <http://www.vrml.org/Specifications/VRML97>.
- Carpenter, M. P. and F. Narin (1973, November-December). Clustering of scientific journals. *Journal of the American Society for Information Science* **24**(6).
- Carrière, J. and R. Kazman (1997). WebQuery: Searching and visualizing the web through connectivity. In *Proceedings of the Sixth International World Wide Web Conference*, pp. 701–711. <http://www.cgl.uwaterloo.ca/Projects/Vanish/webquery-1.html>.
- Celeux, G. (2001). Different points of view for choosing the number of components in a mixture model. In G. Govaert, Janssen, and Limios (Eds.), *Proceedings ASMDA 2001, Compiègne*.
- Celeux, G., C. Biernacki, and G. Govaert (2001). Choosing models in model-based clustering and discriminant analysis. <http://www.inrialpes.fr/is2/people/celeux/asa97.ps>. Working paper available on the Internet.
- Celeux, G. and G. Covaert (1995). Gaussian parsimonious clustering models. *Pattern Recognition* **28**(5): 781–793.
- Chakrabarti, S., B. Dom, S. R. Kumar, P. Raghavan, S. Rajagopalan, A. Tomkins, J. M. Kleinberg, and D. Gibson (1999, June). Hypersearching the web. *Scientific American*: 44–52. <http://www.sciam.com/1999/0699issue/0699raghavan.html>.
- Chance, B., Y. Chen, and C. Cowen (1999, June). Functional activation imaging measured with NIRS. See Rosen, Seitz, and Volkmann (1999), pp. S248. ISSN 1053–8119.
- Charniak, E. (1993). *Statistical Language Learning*. Cambridge, Massachusetts: MIT Press.
- Charniak, E. (1997). Statistical techniques for natural language parsing. *AI Magazine*.

- Chen, A. C. N., F. Å. Nielsen, and L. K. Hansen (1999, June). A 3D brain model of VRML/Talairach for standardized data-sets in neuroimaging of human pain: PET data in headache vs. angina (demo). *NeuroImage* **9**(6): S183. <http://hendrix.imm.dtu.dk/staff/fnielsen/html/ChenAndrew1999A3D.html>.
- Chen, C. (1997, April). Structuring and visualising the world-wide web with generalised similarity analysis. In *Proceedings of the 8th ACM Conference on Hypertext (Hypertext '97)*. ACM. <http://www.brunel.ac.uk/~cssrccc2/papers/ht97.pdf>.
- Chen, C. (1999). Visualising semantic spaces and author co-citation networks in digital libraries. *Information Processing and Management* **35**(3): 401–420. http://www.brunel.ac.uk/~cssrccc2/papers/ip_m/ip_m.pdf.
- Chen, C. and L. Carr (1999, October). Visualizing the evolution of a subject domain: A case study. In *IEEE InfoVis'99*, pp. 449–452. IEEE. <http://www.brunel.ac.uk/~cssrccc2/papers/infvis99.pdf>.
- Chernoff, H. (1973). Using faces to represent points in k -dimensional space graphically. *Journal of the American Statistical Association* **68**: 361–368.
- Chicurel, M. (2000, August). Databasing the brain. *Nature* **406**: 822–825. ISSN 0028-0836. http://www.hirn.uni-duesseldorf.de/rk/Pdf/Nature_BrainDatabasing.pdf. News feature.
- Chittaro, L. (2001). Information visualization and its application to medicine. *Artificial Intelligence in Medicine* **22**(2): 81–88. ISSN 0933-3657. PMID: 11348841. <http://www.elsevier.nl/geom-ng/10/10/70/59/31/27/abstract.html>.
- Chkhenkeli, S., J. Milton, R. K. Erickson, D. M. Frim, J.-P. Spire, and V. L. Towle (1999, June). Alteration of ECoG coherence patterns during seizures. See Rosen, Seitz, and Volkmann (1999), pp. S697. ISSN 1053–8119.
- Christiansen, T. B. (2000, February). Clustering positron emission tomography time series. Master's thesis, Technical University of Denmark, Department of Mathematical Modelling, Lyngby, Denmark. IMM-EKS-2000-2.
- Clare, S. (1997, October). *Functional MRI: Methods and Applications*. Ph. D. thesis, University of Nottingham. <http://www.fmrib.ox.ac.uk/~stuart/thesis/>.
- Cline, H. E., S. Ludke, and W. E. Lorensen (1988, January). Dividing cubes system and method for the display of surface structures contained within the interior region of a solid body. United States Patent No. 4,719,585.
- Cocosco, C. A. and A. C. Evans (2001). Java internet viewer: A WWW tool for remote 3D medical image data visualization and comparison. In *Fourth International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI 2001)*, Utrecht, The Netherlands, 14-17 October 2001.
- Cocosco, C. A., V. Kollokian, R. K.-S. Kwan, and A. C. Evans (1997, May). BrainWeb: Online interface to a 3D MRI simulated brain database. See Friberg, Gjedde, Holm, Lassen, and Nowak (1997). ISSN 1053–8119. http://www.bic.mni.mcgill.ca/users/crisco/HBM97_abs/HBM97_abs.ps.gz.
- Cohen, D. (1968). Magnetoencephalography: Evidence of magnetic fields produced by alpha-rhythm currents. *Science* **161**: 784–786.
- Cohen, J. D. (1995). Highlights: Language- and domain-independent automatic indexing terms for abstracting. *Journal of the American Society for Information Science* **46**(3): 162–174.
- Cohen, M. S. and R. M. DuBois (1999, July). Stability, repeatability, and the expression of signal magnitude in functional magnetic resonance imaging. *Journal of Magnetic Resonance Imaging* **10**(1): 33–40. PMID: 10398975. <http://www3.interscience.wiley.com/cgi-bin/abstract/62500230/START>.
- Cohn, D. and H. Chang (2000). Learning to probabilistically identify authoritative documents. In *Proceedings of the 17th International Conference on Machine Learning*.
- Cohn, D. and T. Hofman (2001). The missing link — a probabilistic model of document content and hypertext connectivity. See Leen, Dietterich, and Tresp (2001). <http://www.cs.cmu.edu/~cohn/papers/nips00.pdf>, <http://nips.djvuzone.org/djvu/nips13/CohnHofmann.djvu>. NIPS-13.
- Cointepas, Y., J.-F. Mangin, L. Garnero, J.-B. Poline, and H. Benali (2001, June). BrainVISA: Software platform for visualization and analysis of multi-modality brain data. *NeuroImage* **13**(6): 98. <http://www.academicpress.com/www/journal/hbm2001/11140.html>.
- Collins, D. L., C. J. Holmes, T. M. Peters, and A. C. Evans (1995). Automatic 3-D model-based neuroanatomical segmentation. *Human Brain Mapping* **3**(3): 190–208.

- Collins, D. L., P. Neelin, T. M. Peters, and A. C. Evans (1994, March/April). Automatic 3D inter-subject registration of MR volumetric data in standardized Talairach space. *Journal of Computer Assisted Tomography* **18**(2): 192–205. PMID: 8126267.
- Collins, D. L., A. P. Zijdenbos, V. Kollokian, J. G. Sled, N. J. Kabani, and C. J. Holmes (1998, June). Design and construction of a realistic digital brain phantom. *IEEE Transactions on Medical Imaging* **17**(3): 463–468. <http://www.bic.mni.mcgill.ca/users/louis/papers/phantom/>.
- Comaniciu, D., P. Meer, D. Foran, and A. Medl (1998). Bimodal system for interactive indexing and retrieval of pathology images. In *Proceedings of the 4th IEEE Workshop on Applications of Computer Vision (WACV'98), Princeton, NJ*, pp. 76–81. ResearchIndex: <http://citeseer.nj.nec.com/comaniciu98bimodal.html>.
- Computational Systems Neuroscience Group (CSN), C. & O. Vogt Brain Research Institute (2001). *Standardized rules for collation, representation and coding of data in the CoCoMac database* (04-05-01 ed.). Düsseldorf, Germany: Computational Systems Neuroscience Group (CSN), C. & O. Vogt Brain Research Institute. http://www.cocomac.org/cocomac_manual.pdf.
- Conradsen, K. (1984a). *En Introduktion til Statistik* (4. ed.). Lyngby, Denmark: IMSOR, DTH.
- Conradsen, K. (1984b). *En Introduktion til Statistik* (Fourth ed.), Volume 2A. Lyngby, Denmark: IMSOR, DTH. In Danish.
- Corbetta, M., F. M. Miezin, G. L. Shulman, and S. E. Petersen (1993, March). A PET study of visuospatial attention. *The Journal of Neuroscience* **13**(3): 1202–1226. ISSN 0270-6474. PMID: 8441008.
- Cordes, D., P. A. Turski, and J. A. Sorensen (2000, November). Compensation of susceptibility-induced signal loss in echo-planar imaging for functional applications. *Magn. Reson. Imaging* **18**(9): 1055–1068. PMID: 11118760.
- Cormack, A. M. (1963). Representation of a function by its line integrals, with some radiological applications. *Journal of Applied Physics* **34**: 2722–2727.
- Cowan, J. D., G. Tesauero, and J. Alspector (Eds.) (1994). *Advances in Neural Information Processing Systems: Proceedings of the 1993 Conference*, San Francisco, CA. Morgan Kaufman Publishers. NIPS-6.
- Cox, D. R. and H. D. Miller (1965). *The Theory of Stochastic Processes*. London, UK: Chapman & Hall. ISBN 0412151707.
- Cox, R. W. (1996, June). AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical Research* **29**(3): 162–173. PMID: 8812068. http://afni.nimh.nih.gov/ssc/afni_paper1.ps.
- Croft, B. (2001). Statistical nature of texts. Class notes, University of Massachusetts, Massachusetts. <http://ciir.cs.umass.edu/cmpsi646/slides-2001/ir2-2001.ps>.
- Cutting, D. R., D. R. Karger, J. O. Penderson, and J. W. Tukey (1992, August). Scatter/gather: A cluster-based approach to browsing large document collections. In *The 15th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 318–329.
- Dachille, F., K. Kreeger, B. Chen, I. Bitter, and A. Kaufman (1998). High-quality volume rendering using texture mapping hardware. In *Proceedings of Eurographics/SIGGRAPH workshop on graphics hardware*, pp. 69–76. ResearchIndex: <http://citeseer.nj.nec.com/dachille98highquality.html>.
- Dale, A. M. and R. L. Buckner (1997). Selective averaging of rapidly presented individual trials using fMRI. *Human Brain Mapping* **5**: 329–340.
- Dale, A. M. and M. I. Sereno (1993). Improved localization of cortical activity by combining EEG and MEG with MRI cortical surface reconstruction: A linear approach. *Journal of Cognitive Neuroscience* **5**: 162–176.
- Damadian, R. (1971). Tumor detection by nuclear magnetic resonance. *Science* **171**: 1151–1153.
- Davatzikos, C. (1997, May). Spatial transformation and registration of brain images using elastically deformable models. *Computer Vision and Image Understanding* **66**(2): 207–222. PMID: 11543561. ResearchIndex: <http://citeseer.nj.nec.com/davatzikos97spatial.html>. Special issue on Medical Imaging.
- Davatzikos, C., H. H. Li, E. Herskovits, and S. M. Resnick (2001, January). Accuracy and sensitivity of detection of activation foci in the brain via statistical parametric mapping: A study using a PET simulator. *NeuroImage* **13**(1): 176–184. PMID: 11133320.

- Davidon, W. C. (1959). Variable metric method for minimization. Technical Report ANL-5990 (revised), AEC Res. Dev.
- Davidon, W. C. (1968). Variance algorithms for minimization. *Computer Journal* **10**: 406–410.
- Davis, K. D., I. Giannoylis, J. Downar, C. L. Kwan, D. J. Mikulis, A. P. Crawley, K. Nicholson, and A. Mailis (2001, June). SI cortex is shut down in chronic pain patients with hysterical anaesthesia. *NeuroImage* **13**(6, part 2): S781. <http://www.apnet.com/www/journal/hbm2001/9764.html>.
- Day, N. E. (1969). Estimating the components of a mixture of normal distributions. *Biometrika* **56**(3): 463–474.
- de Jong, B. M., S. Shipp, B. Skidmore, and R. S. J. Frackowiak (1994, October). The cerebral activity related to the visual perception of forward motion in depth. *Brain* **117**(Pt 5): 1039–1054. PMID: 7953587.
- de Jong, S. (1993). SIMPLS: An alternative approach to partial least squares regression. *Chemometrics and Intelligent Laboratory Systems* **18**: 251–263.
- de Jong, S. and H. A. L. Kiers (1992). Principal covariates regression. *Chemometrics and Intelligent Laboratory Systems* **14**: 155–164.
- de Jong, S. and A. Phatak (1997). Partial least squares regression. pp. 25–36. Society for Industrial and Applied Mathematics: SIAM. ISBN 0–89871–393–5. Proceedings of the Second International Workshop on Total Least Square and Errors-in-Variables Modeling.
- De Luna, X. (1998). An improvement of Akaike's FPE criterion to reduce its variability. *Journal of Time Series Analysis* **19**(4): 457–471.
- Decety, J. and J. Grèzes (1999, May). Neural mechanisms subserving the perception of human action. *Trends in Cognitive Sciences* **3**(5): 172–178.
- Deerwester, S., S. T. Dumais, G. W. Furnas, T. K. Landauer, and R. Harshman (1990, September). Indexing by latent semantic analysis. *Journal of the American Society for Information Science* **41**(6): 391–407. <http://www.si.umich.edu/~furnas/POSTSCRIPTS/LSIJASIS.paper.ps>. ResearchIndex: <http://citeseer.nj.nec.com/deerwester90indexing.html>.
- Deerwester, S. C., S. T. Dumais, G. W. Furnas, R. A. Harshman, T. K. Landauer, K. E. Lochbaum, and L. Q. Streeter (1989, June). Computer information retrieval using latent semantic structure. United States Patent No. 4,839,853. http://www.delphion.com/details?pn=US04839853__.
- Dehaene, S., E. Spelke, P. Pinel, R. Stanescu, and S. Tsivkin (1999, May). Sources of mathematical thinking: behavioral and brain-imaging evidence. *Science* **284**(5416): 970–974. PMID: 10320379.
- Deichmann, R. and R. Turner (2001, June). Compensation of susceptibility induced BOLD sensitivity losses in echo-planar fMRI imaging. *NeuroImage* **13**(6, part 2): S8. <http://www.apnet.com/www/journal/hbm2001/10088.html>.
- DeMers, D. and G. W. Cottrell (1993). Nonlinear dimensionality reduction. See Hanson, Cowan, and Lee Giles (1993), pp. 580–587. NIPS-5.
- Dempster, A. P., N. M. Laird, and D. B. Rubin (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B* **39**: 1–38.
- Descombes, X., F. Kruggel, and D. Y. von Cramon (1998, November). fMRI signal restoration using a spatio-temporal Markov random field preserving transitions. *NeuroImage* **8**(4): 340–349. PMID: 9811552.
- Desmond, J. E., J. M. Sum, A. D. Wagner, J. B. Demb, P. K. Shear, G. H. Glover, J. D. Gabrieli, and M. J. Morrell (1995, December). Functional MRI measurement of language lateralization in wada-tested patients. *Brain* **118**(Pt 6): 1411–1419. PMID: 8595473.
- Devlin, J. T., R. P. Russell, M. H. Davis, C. J. Price, J. Wilson, P. M. Matthews, and L. Tyler (2000a, May). Susceptibility and semantics: Comparing PET and fMRI on a language task. See Fox and Lancaster (2000). ISSN 1053-8119.
- Devlin, J. T., R. P. Russell, M. H. Davis, C. J. Price, J. Wilson, H. E. Moss, P. M. Matthews, and L. K. Tyler (2000b, June). Susceptibility-induced loss of signal: Comparing PET and fMRI on a semantic task. *NeuroImage* **11**(6): 589–600. PMID: 10860788.
- Diamantaras, K. I. and S.-Y. Kung (1994a, November). Cross-correlation neural network models. *IEEE Transaction on Signal Processing* **42**(11): 3218–3223.

- Diamantaras, K. I. and S.-Y. Kung (1994b, September). Multi-layer neural networks for reduced-rank approximation. *IEEE Trans. Neural Networks* **5**(5): 684–697.
- Diamantaras, K. I. and S.-Y. Kung (1996). *Principal Component Neural Networks: Theory and Applications*. Wiley Series on Adaptive and Learning Systems for Signal Processing, Communications, and Control. New York: Wiley. ISBN 0-471-05436-4.
- Diligenti, M., F. M. Coetzee, S. Lawrence, C. L. Giles, and M. Gori (2000). Focused crawling using context graphs. In *Proc. Very Large Databases*.
- Ding, X., T. Masaryk, P. Ruggieri, and J. Tkach (1996). Detection of activation patterns in dynamic functional MRI with a clustering technique. In *Proceedings of the International Society for Magnetic Resonance in Medicine, Fourth Scientific Meeting and Exhibition, New York, New York, USA, April 27 - May 3, 1996*, Volume 3, Berkely, California, pp. 1798. Society of Magnetic Resonance. ISSN 1065-9989.
- Ding, X., J. Tkach, P. Ruggieri, and T. Masaryk (1994, August). Analysis of the time-course functional MRI data with clustering method without use of reference signal. In *Proceedings of the Society of Magnetic Resonance, Second Meeting, August 6–12, 1994, San Francisco, California*, Berkeley, California, pp. 630. Society of Magnetic Resonance. ISSN 1065-9889.
- Djeraba, C. and M. Bouet (1997, November). Digital information retrieval. In F. Golshani and K. Makki (Eds.), *Sixth International Conference on Information and Knowledge Management*, pp. 185–200. ACM: ACM Press.
- Dobelle, W. H. (2000, January/February). Artificial vision for the blind by connecting a television camera to the visual cortex. *ASAIO Journal* **46**(1): 3–9. PMID: 10667705. <http://www.dobelle.com/vision/asaio1.html>.
- Doddi, S., M. V. Marathe, S. S. Ravi, D. S. Taylor, and P. Widmayer (2000). Approximation algorithms for clustering to minimize the sum of diameters. *Nordic Journal of Computing* **7**: 185–203.
- Drury, H. A. and D. C. Van Essen (1997, September). *CARET User's Guide*. St. Louis, Missouri, USA: Washington University School of Medicine.
- Drury, H. A., D. C. Van Essen, and C. H. Anderson (2000, May). SureFit: Software for segmenting the cerebral cortex and generating surface reconstructions. See Fox and Lancaster (2000), pp. S914. ISSN 1053-8119. .
- Drury, H. A., B. West, and D. C. Van Essen (1997, September). *CARETdaemon USER's Guide: Web-based Surface Visualization Software for Atlases of the Cerebral Cortex*. St. Louis, Missouri 63110, USA: Washington University School of Medicine, Department of Anatomy & Neurobiology.
- Duda, R. O., P. E. Hart, and D. G. Stork (2000). *Pattern Classification* (Second ed.), Volume November. New York: John Wiley & Sons, Inc.. ISBN 0-471-05669-3.
- Dumbreck, A. A. and C. W. Smith (1992). 3D TV displays for industrial applications. In *IEE Colloquium on "Stereoscopic Television" (Digest No.173)*, pp. 7/1–4. IEE.
- Duncan, J., R. J. Seitz, J. Kolodny, D. Bor, H. Herzog, A. Ahmed, F. N. Newell, and H. Emslie (2000, July). A neural basis for general intelligence. *Science* **289**(5478): 457–460. <http://www.sciencemag.org/cgi/content/abstract/289/5478/457>.
- Eastwood, J. D. (2000). Dynamic CT brain perfusion imaging. GE Medical Systems. <http://www.gemedicalsystems.com/rad/ct/pdf/perf2.pdf>.
- Ebert, D. S. and C. K. Nicholas (Eds.) (1997, November). *Workshop on New Paradigms in Information Visualization and Manipulation*, New York, New York. The Association of Computing Machinery: ACM Press. ISBN 1581130511.
- Eckardt, C. and G. Young (1939). A principal axis transformation for non-hermitian matrices. *Bull. Amer. Math Soc.* **45**: 118–121.
- Eddy, W. F., M. Fitzgerald, C. R. Genovese, A. Mockus, and D. C. Noll (1996). Functional image analysis software — computational olio. In A. Prat (Ed.), *Proceedings in Computational Statistics*, pp. 39–49. Physica-Verlag.
- Einthoven, W. (1885). *Stereoscopie door kleurverschil*. Doctor's thesis, University of Utrecht.
- Eom, S. B. and R. S. Farris (1996). The contributions of organizational science to the development of decision support systems research specialities. *Journal of the American society for Information Science*: 941–952.
- Erb, M., E. Hülsmann, U. Klose, S. Thesen, and W. Grodd (2001, June). Brain activation mapping of leg movement using fmri with prospective motion correction. *NeuroImage* **13**(6): S9. <http://www.apnet.com/www/journal/hbm2001/11472.html>.

- Ernst, T. and J. Hennig (1994, July). Observation of a fast response in function MR. *Magnetic Resonance in Medicine* **32**(1): 146–149. PMID: 8084231.
- Evans, A. C., D. L. Collins, and C. J. Holmes (1996). Automatic 3D regional MRI segmentation and statistical probability anatomy maps. In R. Myers, V. J. Cunningham, D. L. Bailey, and T. Jones (Eds.), *Quantification of Brain Function Using PET*, Chapter 25, pp. 123–130. San Diego, California: Academic Press. ISBN 0123897602.
- Everitt, B. S. (1979, March). Unresolved problems in cluster analysis. *Biometrics* **35**: 169–181.
- Everitt, B. S. and E. T. Bullmore (1999). Mixture model mapping of brain activation in functional magnetic resonance images. *Human Brain Mapping* **7**(1): 1–14. PMID: 9882086.
- Fabricsius, T., P. Kidmose, and L. K. Hansen (2001). Dynamic components of linear stable mixtures from fractional low order moments. <http://eivind.imm.dtu.dk/staff/tf/>.
- Fadili, M. J., S. Ruan, D. Bloyet, and B. Mazoyer (2000, May). On the number of clusters and the fuzziness index for unsupervised FCA of BOLD fMRI time series. See Fox and Lancaster (2000). ISSN 1053-8119.
- Fahlman, S. E. (1988). Faster-learning variations on back-propagation: An empirical study. See Tourtzky, Hinton, and Sejnowski (1988), pp. 38–51.
- Falkman, G. (2001, May). Information visualisation in clinical odontology: multidimensional analysis and interactive data exploration. *Artificial Intelligence in Medicine* **22**(2): 133–158. PMID: 11348844.
- Farah, M. J. and G. K. Aguirre (1999, May). Imaging visual recognition: PET and fMRI studies of the functional anatomy of human visual recognition. *Trends in Cognitive Sciences* **3**(5): 179–186. PMID: 10322474.
- Feng, D., W. Cai, and R. Fulton (1999). A reliable unbiased parametric imaging algorithm for noisy clinical brain PET data. See The International Society for Cerebral Blood Flow and Metabolism (1999), pp. S806. Brain '99 and BrainPET '99, XIXth International Symposium on Cerebral Blood Flow, Metabolism and Function, IVth International Conference on Quantification of Brain Function with PET.
- Feng, Q. (1997, April). *Algorithms for Drawing Clustered Graphs*. Ph. D. thesis, Department of Computer Science and Software engineering, University of Newcastle. <ftp://ftp.cs.brown.edu/SzpubzSzgd94zSzgd92-93zSzgd93-v2.ps.gz/graph-drawing.ps.gz>. ResearchIndex: <http://citeseer.nj.nec.com/161862.html>.
- Fiacco, A. V. and G. P. McCormick (1968). *Nonlinear programming*. New York: John Wiley.
- Filzmoser, P., R. Baumgartner, and E. Moser (1999, July). A hierarchical clustering method for analyzing functional MR images. *Magnetic Resonance Imaging* **17**(6): 817–826. PMID: 10402588.
- Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics* **7**: 179–188.
- Flake, G. W., S. Lawrence, and C. L. Giles (2000). Efficient identification of web communities. In *Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 150–160. <http://www.neci.nec.com/homepages/flake/kdd2000.ps>.
- Fletcher, P. C., R. J. Dolan, T. Shallice, C. D. Frith, R. S. J. Frackowiak, and K. J. Friston (1996). Is multivariate analysis of PET data more revealing than the univariate approach? Evidence from a study of episodic memory retrieval. *NeuroImage* **3**: 209–215.
- Fletcher, R. (1970). A new approach to variable metric algorithms. *Computer Journal* **13**: 317–322.
- Fletcher, R. (1980). *Practical Methods of optimization*, Volume 1: Unconstrained optimization of Wiley-Interscience Publication. John & Sons.
- Fletcher, R. and M. J. D. Powell (1963). A rapidly convergent descent method for minimization. *Computer Journal* **6**: 163–168.
- Fletcher, R. and C. M. Reeves (1964). Function minimization by conjugate gradient. *Computer Journal* **7**: 149–154.
- Ford, J., F. Makedon, V. Megalooikonomou, L. Shen, T. Steinberg, and A. J. Saykin (2001, June). Spatial comparison of fMRI activation maps for data mining. *NeuroImage* **13**(6): S1302. <http://www.apnet.com/www/journal/hbm2001/11743.html>.
- Forman, S. D., J. D. Cohen, M. Fitzgerald, W. F. Eddy, M. A. Mintun, and D. C. Noll (1995, May). Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): Use of a cluster-size threshold. *Magnetic Resonance in Medicine* **33**(5): 636–647. PMID: 7596267.

- Fortier, J. J. (1966). Simultaneous linear prediction. *Psychometrika* **31**: 369–381.
- Foster, D. and E. George (1994). The risk inflation factor in multiple linear regression. *Ann. Statist.* **22**: 1947–1975.
- Fox, P. T. (1995). Broca's area: motor encoding in somatic space. *Behavioral & Brain Sciences* **18**: 344–345.
- Fox, P. T. (1997). The growth of human brain mapping. *Human Brain Mapping* **5**(2): 1+. ISSN 1065-9471.
- Fox, P. T., A. Huang, L. M. Parsons, J.-H. Xiong, F. Zamarippa, L. Rainey, and J. L. Lancaster (2001, January). Location-probability profiles for the mouth region of human primary motor-sensory cortex: Model validation. *NeuroImage* **13**(1): 196–209. ISSN 1053-8119. PMID: 11133322.
- Fox, P. T., A. Y. Huang, L. M. Parsons, J.-H. Xiong, L. Rainey, and J. L. Lancaster (1999). Functional volumes modeling: Scaling for group size in averaged images. *Human Brain Mapping* **8**(2–3): 143–150. PMID: 10524606. Special Issue: Proceedings of the BrainMap '98 Workshop.
- Fox, P. T., R. J. Ingham, J. C. Ingham, T. B. Hirsch, J. H. Downs, C. Martin, P. Jerabek, T. Glass, and J. L. Lancaster (1996). A PET study of the neural systems of stuttering. *Nature* **382**: 158–162. ISSN 0028-0836.
- Fox, P. T. and J. L. Lancaster (1994, November). Neuroscience on the net. *Science* **266**(5187): 994–996. PMID: 7973682.
- Fox, P. T. and J. L. Lancaster (Eds.) (2000, May). *Sixth International Conference on Functional Mapping of the Human Brain. NeuroImage*, Volume 11. Academic Press. ISSN 1053-8119.
- Fox, P. T., J. L. Lancaster, G. Davis, and S. A. Mikiten (1995). BrainMap : A community database of human functional neuroanatomy. In *First International conference on Functional Mapping of the Human Brain*.
- Fox, P. T., J. L. Lancaster, L. M. Parsons, and J.-H. Xiong (1997a, May). Functional volumes modeling: Metanalytic models for statistical parametric imaging. See Friberg, Gjedde, Holm, Lassen, and Nowak (1997), pp. S397. ISSN 1053–8119.
- Fox, P. T., J. L. Lancaster, L. M. Parsons, J.-H. Xiong, and F. Zamarripa (1997b). Functional volumes modeling: Theory and preliminary assessment. *Human Brain Mapping* **5**(4): 306–311. <http://www3.interscience.wiley.com/cgi-bin/abstract/56435/START>.
- Fox, P. T., F. M. Miezin, J. M. Allman, D. C. Van Essen, and M. E. Raichle (1987). Retinotopic organization of human visual cortex mapped with positron emission tomography. *Journal of Neuroscience* **7**: 919–922.
- Fox, P. T. and M. A. Mintun (1989). Noninvasive functional brain mapping by change-distribution analysis of average PET images of $H_2^{15}O$ tissue activity. *J. Nucl. Med* **30**: 141–149.
- Fox, P. T., L. M. Parsons, and J. L. Lancaster (1998, April). Beyond the single study: functional/location metanalysis in cognitive neuroimaging. *Current Opinion in Neurobiology* **8**(2): 178–187. ISSN 0959-4388.
- Fox, P. T. and M. E. Raichle (1984, May). Stimulus rate dependence of regional cerebral blood flow in human striate cortex, demonstrated by positron emission tomography. *Journal of Neurophysiology* **51**(5): 1109–1120. PMID: 6610024.
- Fox, P. T. and M. E. Raichle (1986). Focal physiological uncoupling of cerebral blood flow and oxidative metabolism during somatosensory stimulation in human subjects. *Proceedings of the National Academy of Sciences USA* **83**: 1140–1144.
- Fox, P. T., M. E. Raichle, M. A. Mintun, and C. Dence (1988). Nonoxidative glucose consumption during focal physiologic neural activation. *Science* **241**: 462–464.
- Frackowiak, R. S. J., K. J. Friston, C. D. Frith, R. Dolan, and J. C. Mazziotta (1997, August). *Human Brain Function*. San Diego, California: Academic Press. ISBN 0122648404.
- Frakes, W. B. (1992). Stemming algorithms. In W. B. Frakes and R. Baez-Yates (Eds.), *Information Retrieval: Data Structures and Algorithms*, pp. 131–160. Prentice-Hall.
- Fraley, C. and A. E. Raftery (1998). How many clusters? which clustering method? answers via model-based cluster analysis. *The Computer Journal* **41**(8): 578–588. <http://lib.stat.cmu.edu/S/mclust/tr329.ps>. ResearchIndex: cite-seer.nj.nec.com/fraley98how.html.
- Franceschini, M. A., V. Toronov, M. E. Filiaci, M. Wolf, A. Michalos, E. Gratton, and S. Fantini (2000, May). Real-time video of cerebral hemodynamics in the human brain using non-invasive optical imaging. See Fox and Lancaster (2000), pp. S454. ISSN 1053-8119.

- Fredriksson, J., P. E. Roland, P. Svensson, K. Amunts, C. Cavada, R. Hari, A. Cowey, F. Crivello, S. Geyer, G. Kostopoulos, B. Mazoyer, D. Popplewell, A. Schleicher, T. Schormann, M. Seppa, H. Uylings, K. de Vos, and K. Zilles (2000, May). The European computerised human brain database. See Fox and Lancaster (2000), pp. S906. ISSN 1053-8119. <http://www.academicpress.com/www/journal/hbm2000/6278.html>.
- Freeman, L. C. (2000, February). Visualizing social networks. *Journal of Social Structure* 1(1). <http://www.library.cmu.edu:7850/JoSS/article.html>.
- French, J. C. and C. L. Villes (1996, December). Exploiting coauthorship to infer topicality in a digital library of computer science technical reports. Technical Report CS-96-20, Department of Computer Science, University of Virginia. <ftp://ftp.cs.virginia.edu/pub/techreports/CS-96-20.ps.Z>.
- Friberg, L., A. Gjedde, S. Holm, N. A. Lassen, and M. Nowak (Eds.) (1997, May). *Third International Conference on Functional Mapping of the Human Brain, NeuroImage*, Volume 5. Academic Press. ISSN 1053-8119.
- Friedrich, R., A. Fuchs, and H. Haken (1991). Modelling of spatio-temporal EEG patterns. In I. D. I and A. V. Holden (Eds.), *Mathematical approaches to brain functioning diagnostics*. New York: Manchester University Press.
- Friman, O., M. Borga, P. Lundberg, and H. Knutsson (2001a, March). Canonical correlation as a tool in functional MRI data analysis. In *SSAB Symposium 2001, Norrköping, Sweden*. Swedish Society for Automated Image Analysis. <http://media-matlab1.itn.liu.se/papers/frim/Friman.pdf>.
- Friman, O., J. Cedefamn, P. Lundberg, M. Borga, and H. Knutsson (2001b, February). Detection of neural activity in functional MRI using canonical correlation analysis. *Magnetic Resonance in Medicine* 45(2): 323-330. PMID: 11180440.
- Friman, O., P. Lundberg, J. Cedefamn, M. Borga, and H. Knutsson (2001c, April). Increased detection sensitivity in fMRI by adaptive filtering. In *Proceedings of ISMRM, Glasgow, Scotland*. The International Society for Magnetic Resonance in Medicine. ftp://ftp.imt.liu.se/pub/bildb/Papers/poster_OF_a4.pdf.
- Friston, K. J. (1997a, May). Basic concepts and overview. In *SPMcourse, Short course notes*, Chapter 1. Institute of Neurology, Wellcome Department of Cognitive Neurology. <http://www.fil.ion.ucl.ac.uk/spm/course/notes.html>.
- Friston, K. J. (1997b, May). Eigenimages and multivariate analysis. In *SPMcourse, Short course notes*, Chapter 5. Institute of Neurology, Wellcome Department of Cognitive Neurology.
- Friston, K. J. (1997c). Testing for anatomically specified regional effects. *Human Brain Mapping* 5(2): 133-136. PMID: 10096418. <http://www3.interscience.wiley.com/cgi-bin/abstract/56419/START>.
- Friston, K. J. (1999a, May). *spm_hrf.m*. Included in the SPM distribution. Version 2.7.
- Friston, K. J. (1999b, November). *spm_mip.m*. Included in the SPM distribution. Version 2.5.
- Friston, K. J. (2001, May). Re: MANCOVA/CVA paper. SPM mailing list.
- Friston, K. J., J. Ashburner, C. D. Frith, J.-B. Poline, J. D. Heather, and R. S. J. Frackowiak (1995a). Spatial registration and normalization of images. *Human Brain Mapping* 2: 165-189. http://www.fil.ion.ucl.ac.uk/spm/papers/SPM_1/.
- Friston, K. J., P. Fletcher, O. Josephs, A. P. Holmes, M. D. Rugg, and R. Turner (1998a). Event-related fMRI: characterising differential responses. *NeuroImage* 7: 30-40.
- Friston, K. J., C. D. Frith, P. Fletcher, P. F. Liddle, and R. S. J. Frackowiak (1996a, March-April). Functional topography: multidimensional scaling and functional connectivity in the brain. *Cerebral Cortex* 6(2): 156-164. PMID: 8670646. http://www.fil.ion.ucl.ac.uk/spm/papers/PET_2/.
- Friston, K. J., C. D. Frith, R. S. J. Frackowiak, and R. Turner (1995b, June). Characterizing dynamic brain responses with fMRI: A multivariate approach. *NeuroImage* 2(2): 166-172. http://www.fil.ion.ucl.ac.uk/spm/papers/fMRI_5, <http://www.idealibrary.com/links/doi/10.1006/nimg.1995.1019/pdf>.
- Friston, K. J., C. D. Frith, P. F. Liddle, R. J. Dolan, A. A. Lammertsmaa, and R. S. J. Frackowiak (1990, November). The relationship between global and local changes in PET scans. *Journal of Cerebral Blood Flow and Metabolism* 13(6): 1038-1040. PMID: 2347879.
- Friston, K. J., C. D. Frith, P. F. Liddle, and R. S. J. Frackowiak (1991, July). Comparing functional (PET) images: the assessment of significant change. *Journal of Cerebral Blood Flow and Metabolism* 11(4): 690-699. PMID: 2050758.

- Friston, K. J., C. D. Frith, P. F. Liddle, and R. S. J. Frackowiak (1993, January). Functional connectivity: The principal-component analysis of large (PET) data sets. *Journal of Cerebral Blood Flow and Metabolism* **13**(1): 5–14. PMID: 8417010.
- Friston, K. J., A. P. Holmes, J.-B. Poline, P. J. Grasby, S. C. R. Williams, R. S. J. Frackowiak, and R. Turner (1995c, March). Analysis of fMRI time series revisited. *NeuroImage* **2**(1): 45–53. PMID: 9343589. <http://www.idealibrary.com/links/doi/10.1006/nimg.1995.1007/>, http://www.fil.ion.ucl.ac.uk/spm/papers/fMRI_3/.
- Friston, K. J., A. P. Holmes, and K. J. Worsley (1999a, July). How many subjects constitute a study? *NeuroImage* **10**(1): 1–5. ISSN 1053-8119. PMID: 10385576. <http://www.idealibrary.com/links/citation/1053-8119/10/1>.
- Friston, K. J., A. P. Holmes, K. J. Worsley, J.-B. Poline, C. D. Frith, and R. S. J. Frackowiak (1995d). Statistical parametric maps in functional imaging: A general linear approach. *Human Brain Mapping* **2**: 189–210. http://www.fil.ion.ucl.ac.uk/spm/papers/SPM_3/.
- Friston, K. J., P. Jezzard, and R. Turner (1994a). The analysis of functional MRI time-series. *Human Brain Mapping* **1**: 153–171. http://www.fil.ion.ucl.ac.uk/spm/papers/fMRI_2/.
- Friston, K. J., O. Josephs, G. Rees, and R. Turner (1998b). Nonlinear event-related responses in fMRI. *Magnetic Resonance in Medicine* **39**: 41–52.
- Friston, K. J., O. Josephs, E. Zarahn, A. P. Holmes, S. Rouquette, and J.-B. Poline (2000a, August). To smooth or not to smooth? bias and efficiency in fmri time-series analysis. *NeuroImage* **12**(2): 196–208. PMID: 10913325.
- Friston, K. J., A. Mechelli, R. Turner, and C. J. Price (2000b). Nonlinear responses in fMRI: The balloon model, volterra kernel, and other hemodynamics. *NeuroImage* **12**: 466–477.
- Friston, K. J., J. Philips, D. Chawla, and C. Büchel (1999b). Revealing interaction among brain systems with nonlinear PCA. *Human Brain Mapping* **8**(2–3): 92–97. PMID: 10524598.
- Friston, K. J., J. Philips, D. Chawla, and C. Büchel (2000c). Nonlinear PCA: Characterizing interactions between modes of brain activity. *Phil. Trans. R. Soc. Lon. B* **355**(1393): 135–146. PMID: 10703049.
- Friston, K. J., J.-B. Poline, A. P. Holmes, C. D. Frith, and R. S. J. Frackowiak (1996b). A multivariate analysis of PET activation studies. *Human Brain Mapping* **4**(2): 140–151. <http://www3.interscience.wiley.com/cgi-bin/abstract/73070/START>, http://www.fil.ion.ucl.ac.uk/spm/papers/PET_1/.
- Friston, K. J., C. J. Price, P. Fletcher, C. Moore, R. S. J. Frackowiak, and R. J. Dolan (1996c, October). The trouble with cognitive subtraction. *NeuroImage* **4**(2): 97–104. PMID: 9345501.
- Friston, K. J., S. Williams, R. S. J. Frackowiak, J. C. Mazziotta, and A. C. Evans (1996d). Movement-related effects in fMRI timeseries. *Magnetic Resonance in Medicine* **35**: 346–355.
- Friston, K. J., K. J. Worsley, R. S. J. Frackowiak, and R. Turner (1994b). Assessing the significance of focal activations using their spatial extent. *Human Brain Mapping* **1**: 210–220.
- Frutiger, S. A., S. C. Strother, J. R. Anderson, J. J. Sidtis, J. B. Arnold, and D. A. Rottenberg (2000, November). Multivariate predictive relationship between kinematic and functional activation patterns in a PET study of visuomotor learning. *NeuroImage* **12**(5): 515–527. PMID: 11034859. <http://www.idealibrary.com/links/citation/1053-8119/12/515>.
- Fu, W. J. (1998). Penalized regressions: The bridge versus the lasso. *Journal of Computational and Graphical Statistics* **7**(3): 397–416.
- Fuchs, A., J. A. S. Kelso, and H. Haken (1992). Phase transitions in the human brain: spatial mode dynamics. *Int. J. of Bifurcation and Chaos* **2**: 917–939.
- Fuchs, H., Z. M. Kedem, and S. P. Uelson (1977, October). Optimal surface reconstruction from planar contours. *Communications of the ACM* **20**(10): 693–702.
- Fujimura, K. and E. Kuo (1999). Shape reconstruction from contours using isotopic deformation. *Graphical models and image processing: GMIP* **61**(3): 127–147. ResearchIndex: <http://www.researchindex.com/370999.html>.
- Gade, A. (1997). *Hjerneprocesser. Kognition og neurovidenskab*. Copenhagen, Denmark: Frydenlund. In Danish.
- Gale, W. A. and G. Sampson (1995). Good-Turing frequency estimation without tears. *Journal of Quantitative Linguistics* **2**: 217–237.

- Gansner, E. R. and S. C. North (2000). An open graph visualization system and its applications to software engineering. *Software — Practice and Experience* **30**(11): 1203–1234. ISSN 00380644. <http://www.research.att.com/sw/tools/graphviz/GN99.pdf>.
- Gansner, E. R., S. C. North, and K. P. Vo (1988, November). Dag: A program that draws directed graphs. *Software — Practice and Experience* **18**(11): 1047–1062. <ftp://ftp.cs.utexas.edu/pub/code2/kleyn/graphs/AttDag/dagdoc.ps>. ResearchIndex: <http://citeseer.nj.nec.com/gansner89dag.html>.
- Garde, E., E. L. Mortensen, K. Krabbe, E. Rostrup, and H. B. W. Larsson (2000). Relation between age-related decline in intelligence and cerebral white-matter hyperintensities in healthy octogenarians: a longitudinal study. *Lancet* **356**: 628–634. PMID: 10968435.
- Gardner, D., K. H. Knuth, M. Abato, S. M. Erde, T. White, R. DeBellis, and E. P. Gardner (2001, January/February). Common data model for neuroscience data and data model exchange. *Journal of the American Medical Informatics Association* **8**(1): 17–33. <http://www.jamia.org/cgi/content/abstract/8/1/17>.
- Garfield, E. (1972, November). Citation analysis as a tool in journal evaluation. *Science* **178**(4060): 471–479. PMID: 5079701. <http://www.garfield.library.upenn.edu/essays/V1p527y1962-73.pdf>.
- Garfield, E. (1986, October). Towards scientography. In *Essays of an Information Scientist*, Volume 9, pp. 324. <http://www.garfield.library.upenn.edu/essays/v9p324y1986.pdf>.
- Garfield, E. (1990, February). The most-cited papers of all time, SCI 1945–1988. part 1a. the *sci* top 100 — will the Lowry method ever be obliterated. *Current Contents* (7): 3–14. <http://www.garfield.library.upenn.edu/essays/v13p045y1990.pdf>.
- Garfield, E. (1996, October). Citation indexes for retrieval and research evaluation. <http://www.garfield.library.upenn.edu/papers/ciretreseval-capri.html>. Presentation (Unpublished), at the G7 Consensus Conference on Theory and Practice of Research Assessment in Capri, Italy.
- Gaser, C., I. Nenadic, B. R. Buchsbaum, E. A. Hazlett, and M. S. Buchsbaum (2001, June). Deformation-based morphometry and its relation to convention volumetry of brain lateral ventricles in MRI. *NeuroImage* **13**(6): 1140–1145.
- Geiger, B. (1993, December). Three-dimensional modeling of human organs and its application to diagnosis and surgical planning. Technical Report 2105, Institut National de Recherche en Informatique et Automatique, 06902 Sophia Antipolis, France. <ftp://ftp-sop.inria.fr/prisme/NUAGES/Nuages/RR2105.ps.gz>.
- Geisser, S. and S. W. Greenhouse (1958). An extension of Box's results on the use of the F distribution in multivariate analysis. *The Annals of Mathematical Statistics* **29**(885): 885–891.
- Geman, S., E. Bienenstock, and R. Doursat (1992). Neural networks and the bias/variance dilemma. *Neural Computation* **4**(1): 1–58.
- Genovese, C. R., D. C. Noll, and W. F. Eddy (1997). Estimating test-retest reliability in functional MR imaging I: Statistical methodology. *Magnetic Resonance in Medicine* **38**: 497–507.
- George, J., D. Rector, C. Ho, K. Albright, C. Smith, and A. Hielscher (2000, May). Time-resolved photon migration tomography and spectroscopy with RULLI. See Fox and Lancaster (2000), pp. S456. ISSN 1053-8119.
- George, M. S., T. A. Ketter, B. A. Parekh, D. S. Gill, T. Huggins, L. Marangell, P. J. Pazaglia, and R. Post (1994). Spatial ability in affective illness: differences in regional brain activation during a spatial matching task ($H_2^{15}O$ PET). *Neuropsychiatry, Neuropsychology, and Behavioral Neurology* **7**(3): 143–153.
- Gering, D. T. (1999). A system for surgical planning and guidance using image fusion and interventional MR. Master's thesis, MIT.
- Gering, D. T., A. Nabavi, R. Kikinis, W. E. L. Grimson, N. Hata, P. Everett, F. Jolesz, and W. M. Wells (1999, September). An integrated visualization system for surgical planning and guidance using image fusion and interventional imaging. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, Cambridge, England. <http://www.slicer.org>.
- Gerstein, G. L. and D. H. Perkel (1972). Mutual temporal relationships among neuronal spike trains: Statistical techniques for display and analysis. *Biophysical Journal* **12**: 453–473.
- Geyer, S., A. Schleicher, and K. Zilles (1997, July). The somatosensory cortex of human: cytoarchitecture and regional distributions of receptor-binding sites. *NeuroImage* **6**(1): 27–45.

- Gibson, D., J. Kleinberg, and P. Raghavan (1998, June). Inferring web communities from link topology. In *Proceedings of the ninth ACM conference on Hypertext and hypermedia: links, objects, time and space-structure in hypermedia systems*, pp. 225–234. ACM.
- Giles, C. L., K. Bollacker, and S. Lawrence (1998, June). CiteSeer: An automatic citation indexing system. In I. Witten, R. Akscyn, and F. M. S. III (Eds.), *Digital Libraries 98 - The Third ACM Conference on Digital Libraries*, Pittsburgh, PA, pp. 89–98. ACM Press. ISBN 0897919653.
- Glantz, S. A. and L. A. Bero (1994, July). Inappropriate and appropriate selection of “peers” in grant review. *JAMA: The journal of the American Medical Association* **272**(2): 114–116. PMID: 8015118. http://www.ama-assn.org/public/peer/7_13_94/pv3105x.htm. The Second International Congress on Peer Review in Biomedical Publication.
- Glaser, D. E., W. D. Penny, R. N. A. Henson, M. D. Rugg, and K. J. Friston (2001, June). Correcting for non-sphericity in imaging data using classical and Bayesian approaches. *NeuroImage* **13**(6 (part 2)): S127. <http://www.apnet.com/www/journal/hbm2001/11042.html>.
- Glover, E. J., S. Lawrence, W. P. Birmingham, and C. L. Giles (1999, November). Architecture of a metasearch engine that supports user information needs. In *Eighth International Conference on Information and Knowledge Management (CIKM 99)*.
- Glover, G. H. (1999). Deconvolution of impulse response in event-related BOLD fMRI. *NeuroImage* **9**: 416–429. PMID: 10191170. <http://www.idealibrary.com/links/citation/1053-8119/9/416>.
- Gokcay, D., C. M. Mohr, B. Crosson, C. M. Leonard, and J. A. Bobholz (1999, December). LOFA: software for individualized localization of functional mri activity. *NeuroImage* **10**(6): 749–755. PMID: 10600420.
- Gold, S., B. Christian, S. Arndt, G. Zeien, T. Cizadlo, D. L. Johnson, M. Flaum, and N. C. Andreasen (1998). Functional MRI statistical software packages: A comparative analysis. *Human Brain Mapping* **6**(2): 73–84. PMID: 9673664. <http://www3.interscience.wiley.com/cgi-bin/fulltext?ID=38759&PLACEBO=IE.pdf>.
- Goldfarb, D. (1970). A family of variable metric methods derived by variational means. *Maths. Comput.* **24**: 23–26.
- Gollob, H. F. (1968). A statistical model which combines features of factor analytic and analysis of variance techniques. *Psychometrika* **33**: 73–117.
- Golub, G. H. and C. F. Van Loan (1996). *Matrix Computation* (Third ed.). John Hopkins Studies in the Mathematical Sciences. Baltimore: Johns Hopkins University Press.
- Good, I. J. (1953). The population frequencies of species and the estimation of population parameters. *Biometrika* **40**(3 and 4): 237–264.
- Goodhill, G. J., M. W. Simmen, and D. J. Willshaw (1995). An evaluation of the use of multidimensional scaling for understanding brain connectivity. *Philosophical Transaction of the Royal Society, London, Series B* **348**: 265–280. http://www.giccs.georgetown.edu/geoff/pub/goodhill_simmen_willshaw_94.ps.Z. ResearchIndex: <http://citeseer.nj.nec.com/goodhill94evaluation.html>.
- Goualher, G. L., E. Procyk, D. L. Collins, R. Venugopal, C. Barillot, and A. C. Evans (1999, March). Automated extraction and variability analysis of sulcal neuroanatomy. *IEEE Transactions on Medical Imaging* **18**(3): 206–217. PMID: 10363699.
- Gouraud, H. (1971, June). Continuous shading of curved surfaces. *IEEE Transactions on Computers* **20**(6): 623–629.
- Goutte, C. and L. K. Hansen (1997). Regularization with a pruning prior. *Neural Networks* **10**(6): 1053–1059.
- Goutte, C., L. K. Hansen, M. G. Liptrot, and E. Rostrup (1999a, September). Feature space clustering for fMRI meta-analysis. Technical Report IMM-REP-1999-13, Department of Mathematical Modelling, Technical University of Denmark, Lyngby, Denmark.
- Goutte, C., L. K. Hansen, M. G. Liptrot, and E. Rostrup (2001, July). Feature-space clustering for fmri meta-analysis. *Human Brain Mapping* **13**(3): 165–183. <http://www3.interscience.wiley.com/cgi-bin/abstract/82002382/START>.
- Goutte, C. and J. Larsen (1998). Optimal cross-validation split ratio: Experimental investigation. In *Proceedings of ICANN98, Skövde, Sweden, September 1998*. <http://eivind.imm.dtu.dk/publications/1998/goutte.icann98.ps.gz>.
- Goutte, C., F. Å. Nielsen, and L. K. Hansen (2000, December). Modeling the haemodynamic response in fMRI using smooth FIR filters. *IEEE Transactions on Medical Imaging* **19**(12): 1188–1201.

- Goutte, C., F. Å. Nielsen, C. Svarer, E. Rostrup, and L. K. Hansen (1998, May). Space-time analysis of fMRI by feature space clustering. See Paus, Gjedde, and Evans (1998), pp. S610. ISSN 1053–8119. <http://eivind.imm.dtu.dk/publications/1998/Goutte1998Feature.ps.gz>.
- Goutte, C. and E. Rostrup (1999, June). Assessing fMRI activation using maximum convoluted correlation. See Rosen, Seitz, and Volkman (1999), pp. S19. ISSN 1053–8119.
- Goutte, C., P. Toft, E. Rostrup, F. Å. Nielsen, and L. K. Hansen (1999b, March). On clustering fMRI time series. *NeuroImage* **9**(3): 298–310.
- Gower, J. C. (1966). Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika* **53**: 325–328.
- Gower, J. C. (1975). Goodness-of-fit criteria of classification and other patterned structures. In *Procedure of the 8th International Conference on Numerical Taxonomy*, pp. 38–62. W. H. Freeman and Co.
- Grafton, S. T., R. P. Woods, and J. C. Mazziotta (1993). Within-arm somatotopy in human motor areas determined by positron emission tomography imaging of cerebral blood flow. *Experimental Brain Research* **95**(1): 172–176. PMID: 8405250.
- Grant, J., S. Burden, and G. Breen (1997, December). No evidence of sexism in peer review. *Nature* **390**(6659): 438. ISSN 0028-0836.
- Graves, L., A. Pack, and T. Abel (2001, 243). Sleep and memory: a molecular perspective. *Trends in Neuroscience* **24**(4): 237–. PMID: 11250009.
- Grefenstette, G. and P. Tapanainen (1994). What is a word, what is a sentence? Problems of tokenization. In *Proceedings of the 3rd International Conference on Computational Lexicography (COMPLEXt'94)*, Research Institute for Linguistic, Hungarian Academy of Sciences, Budapest, pp. 79–87.
- Greitz, T., C. Bohm, S. Holte, and L. Eriksson (1991, January-February). A computerized brain atlas: construction, anatomical content, and some applications. *Journal of Computer Assisted Tomography* **15**(1): 26–38. PMID: 1987199.
- Grethe, J. S., J. D. Van Horn, J. B. Woodward, S. Inati, P. J. Kostelec, J. A. Aslam, D. Rockmore, D. Rus, and M. S. Gazzaniga (2001, June). The fMRI data center: An introduction. *NeuroImage* **13**(6): S135. <http://www.apnet.com/www/journal/hbm2001/11329.html>.
- Grinvald, A., L. B. Cohen, S. Leshner, and M. B. Boyle (1981). Simultaneous optical monitoring of activity of many neurons in invertebrate ganglia using a 124 element photodiode array. *Journal of Neurophysiology* **45**: 829–840.
- Grinvald, A., E. Lieke, R. D. Frostig, C. D. Sperling, and T. N. Wiesel (1986, November). Functional architecture of cortex revealed by optical imaging of intrinsic signals. *Nature* **324**(6095): 361–364. PMID: 3785405.
- Grinvald, A., B. M. Salzberg, and L. B. Cohen (1977, July). Simultaneous recording from several neurons in an invertebrate central nervous system. *Nature* **268**(5616): 140–142. PMID: 593306.
- Guelich, S., S. Gundavaram, and G. Birznies (2000, July). *CGI Programming with Perl* (Second ed.). Sebastopol, CA: O'Reilly. ISBN 1-56592-419-3. See <http://www.oreilly.com/catalog/cgi2/errata/> for errata.
- Gunderson, B. K. and R. J. Muirhead (1997, July). On estimating the dimensionality in canonical correlation analysis. *Journal of Multivariate Analysis* **62**(1): 121–136.
- Gunn, R. N., J. Ashburner, J. Aston, and V. J. Cunningham (1999). Analysis of functional imaging datasets via functional segmentation. See The International Society for Cerebral Blood Flow and Metabolism (1999), pp. S790. Brain '99 and BrainPET '99, XIXth International Symposium on Cerebral Blood Flow, Metabolism and Function, IVth International Conference on Quantification of Brain Functional with PET.
- Gusnard, D., E. Akbudak, G. Shulman, and M. E. Raichle (2001, June). Role of medial prefrontal cortex in a default mode of brain function. *NeuroImage* **13**(6, part 2): S414. <http://www.apnet.com/www/journal/hbm2001/10683.html>.
- Hagen, T., K. Bartylla, and U. Piepgras (1999). Correlation of regional cerebral blood flow measured by stable xenon CT and perfusion MRI. *Journal of Computer Assisted Tomography* **23**(2): 257–264.
- Hahn, E. L. (1950). Spin echoes. *Physical Review* **80**(4): 580–594.
- Hall, P. A. V. and G. R. Dowling (1980). Approximate string matching. *Computing Surveys* **12**(4): 381–402.

- Hannan, E. J. and B. G. Quinn (1979). The determination of the order of an autoregression. *Journal of the Royal Statistical Society, Series B* **41**(2): 190–195.
- Hansen, L. K. and J. Larsen (1998, June). Source separation in short image sequences using delayed correlation. In P. Dalsgaard and S. H. Jensen (Eds.), *Proceedings of NORSIG'98, Vigsø, Denmark*, pp. 253–256. ISBN 87-985750-8-2. <http://eivind.imm.dtu.dk/publications/1998/hansen.norsig98.ps.Z>. ResearchIndex: <http://citeseer.nj.nec.com/185061.html>.
- Hansen, L. K., J. Larsen, and T. Kolenda (2000a, August). On independent component analysis for multimedia signals. In S.-Y. K. Ling Guan and J. Larsen (Eds.), *Multimedia Image and Video Processing*, Chapter 7, pp. 175–199. CRC Press. ISBN 0849334926.
- Hansen, L. K., J. Larsen, and T. Kolenda (2001a, May). Blind detection of independent dynamic components. In *Proceedings of ICASSP'2001*. <http://eivind.imm.dtu.dk/publications/2001/hansen.icassp2001.pdf>.
- Hansen, L. K., J. Larsen, F. Å. Nielsen, S. C. Strother, E. Rostrup, R. Savoy, C. Svarer, and O. B. Paulson (1999a, May). Generalizable patterns in neuroimaging: How many principal components? *NeuroImage* **9**(5): 534–544.
- Hansen, L. K., N. J. S. Mørch, and F. Å. Nielsen (1998). Neural networks in functional neuroimaging. In *NORSIG'98, IEEE Nordic Signal Processing Symposium, Aalborg University*, pp. 1–8.
- Hansen, L. K., F. Å. Nielsen, M. G. Liptrot, C. Goutte, S. C. Strother, N. Lange, A. Gade, D. A. Rottenberg, and O. B. Paulson (2000b, May). lyngby 2.0 – a modeler's matlab toolbox for spatio-temporal analysis of functional neuroimages. See Fox and Lancaster (2000), pp. S917. ISSN 1053-8119. <http://www.apnet.com/www/journal/hbm2000/6990.html>.
- Hansen, L. K., F. Å. Nielsen, M. G. Liptrot, S. C. Strother, N. Lange, A. Gade, D. A. Rottenberg, and O. B. Paulson (2001b, June). “lyngby” a matlab toolbox for fMRI analysis: Results of a user questionnaire. *NeuroImage* **13**(6): S145. <http://www.apnet.com/www/journal/hbm2001/11156.html>.
- Hansen, L. K., F. Å. Nielsen, S. C. Strother, and N. Lange (2000c, June). Consensus inference in neuroimaging. *NeuroImage* **13**(6): 1212–1218. PMID: 11352627. <http://www.idealibrary.com/links/citation/1053-8119/13/1212>.
- Hansen, L. K., F. Å. Nielsen, P. Toft, M. G. Liptrot, C. Goutte, S. C. Strother, N. Lange, A. Gade, D. A. Rottenberg, and O. B. Paulson (1999b, June). “lyngby” — a modeler's Matlab toolbox for spatio-temporal analysis of functional neuroimages. See Rosen, Seitz, and Volkmann (1999), pp. S241. ISSN 1053-8119. <http://eivind.imm.dtu.dk/publications/1999/hansen.hbm99.ps.gz>.
- Hansen, L. K., F. Å. Nielsen, P. Toft, S. C. Strother, N. Lange, N. J. S. Mørch, C. Svarer, O. B. Paulson, R. Savoy, B. R. Rosen, E. Rostrup, and P. Born (1997, May). How many principal components? See Friberg, Gjedde, Holm, Lassen, and Nowak (1997), pp. S474. ISSN 1053-8119. <http://eivind.imm.dtu.dk/publications/1997/HBM97.principal.poster474.ps.gz>.
- Hansen, L. K. and C. E. Rasmussen (1994). Pruning from adaptive regularization. *Neural Computation* **6**(6): 1222–1231. <http://eivind.imm.dtu.dk>.
- Hansen, L. K., C. E. Rasmussen, C. Svarer, and J. Larsen (1994). Adaptive regularization. See Vlontzos, Hwang, and Wilson (1994), pp. 78–87. <http://eivind.imm.dtu.dk>.
- Hansen, L. K., S. Sigurðsson, T. Kolenda, F. Å. Nielsen, U. Kjems, and J. Larsen (2000d, June). Modeling text with generalizable Gaussian mixtures. In *Proceedings of ICASSP'2000*. Turkey. <http://eivind.imm.dtu.dk/publications/1999/hansen.icassp2000.ps.gz>. ResearchIndex: <http://citeseer.nj.nec.com/hansen99modeling.html>.
- Hansen, P. C. (1996). *Rank-Deficient and Discrete Ill-Posed Problems*. Lyngby, Denmark: Polyteknisk Forlag. ISBN 87-502-0784-9. Doctoral Dissertation.
- Hansen, P. S., N. L. Sørensen, and S. E. B. Sørensen (2000e, April). Method of identifying a pair of colored optical filters for viewing stereogram by measuring spectral transmission characteristics and calculating corresponding coeffs of color matching functions for first and second filters. World patent No. WO200023845.
- Hanson, S. J., J. D. Cowan, and C. Lee Giles (Eds.) (1993). *Advances in Neural Information Processing Systems: Proceedings of the 1992 Conference*, San Mateo, CA. Morgan Kaufmann Publishers. NIPS-5.
- Hartigan, J. A. (1975). *Clustering Algorithms*. Wiley series in probability and mathematical statistics. John Wiley & Sons. ISBN 0-471-35645-X.
- Hartvig, N. V. (1999). A stochastic geometry model for fMRI data. Research report 410, Department of Theoretical Statistics, University of Aarhus, Aarhus Denmark. <http://home.imf.au.dk/vaeuver/thesis/paperI.ps.gz>. ResearchIndex: <http://citeseer.nj.nec.com/hartvig00spatial.html>.

- Hartvig, N. V. (2000, July). *Parametric Modelling of Functional Magnetic Resonance Imaging Data*. Ph. D. thesis, Department of Theoretical Statistics, University of Aarhus, Århus, Denmark. <http://home.imf.au.dk/vaeuver/thesis/>.
- Hartvig, N. V. and J. L. Jensen (2000, December). Spatial mixture modeling of fMRI data. *Human Brain Mapping* **11**(4): 233–248. PMID: 11144753.
- Hassibi, B., D. Stock, and G. Wolff (1992). Optimal brain surgeon and general network pruning. *Neural Computation* (4): 1–8.
- Hassibi, B. and D. G. Stock (1993). Second order derivatives for network pruning: Optimal brain surgeon. See Hanson, Cowan, and Lee Giles (1993), pp. 164–171. NIPS-5.
- Hasson, R. and S. J. Swithenby (2000). The bayesian power imaging (bpi) method for magnetic source imaging. In J. Nenonen, R. Ilmoniemi, and T. Katila (Eds.), *Biomag2000, Proceedings of the 12th International Conference on Biomagnetism*, Espoo, Finland, pp. 701–704. Helsinki University of Technology. <http://biomag2000.hut.fi/papers/0701.pdf>.
- Hastie, T. and R. J. Tibshirani (1996). Discriminant analysis by Gaussian mixtures. *Journal of the Royal Statistical Society, Series B, Methodology* **58**(1): 155–176. ResearchIndex: <http://citeseer.nj.nec.com/hastie96discriminant.html>.
- Hastie, T. J. and R. J. Tibshirani (1990). *Generalized Additive Models*. London: Chapman & Hall.
- Haxby, J. V., L. Petit, L. G. Ungerleider, and S. M. Courtney (2000, May). Distinguishing the functional roles of multiple regions in distributed neural systems for visual working memory. *NeuroImage* **11**(5): 380–391.
- Haykin, S. (1994). *Neural Networks*. New York: Macmillan College Publishing Company. ISBN 0-02-352761-7.
- Haznedar, M. M., M. S. Buchsbaum, T. C. Wei, P. R. Hof, C. Cartwright, C. A. Bienstock, and E. Hollander (2000, December). Limbic circuitry in patients with autism spectrum disorders studied with positron emission tomography and magnetic resonance imaging. *American Journal of Psychiatry* **157**(12): 1994–2001. PMID: 11097966.
- Heimer, L. (1994). *The human brain and spinal cord: functional neuroanatomy and dissection guide* (Second ed.). New York: Springer-Verlag. ISBN 0-387-94227-0.
- Henson, R. N. A. and M. D. Rugg (2001, June). Effects of stimulus repetition on latency of BOLD impulse response. *NeuroImage* **13**(6): S683. <http://www.apnet.com/www/journal/hbm2001/9690.html>.
- Henzinger, M. (2000, September). Link analysis in web information retrieval. *Bulletin of the Technical Committee on Data Engineering* **23**(3): 3–8. <http://www.research.microsoft.com/research/db/debull/A00sept/henzinge.ps>.
- Herholz, K., A. Thiel, K. Wienhard, U. Pietrzyk, H.-M. von Stockhausen, H. Karbe, J. Kessler, T. Bruckbauer, M. Halber, and W.-D. Weiss (1996, June). Individual functional anatomy of verb generation. *NeuroImage* **3**(3, Part 1 of 2): 185–194.
- Herskovits, E. H. (2000a, December). An architecture for a brain-image database. *Methods of Information in Medicine* **39**(4–5): 291–297. PMID: 11191696.
- Herskovits, E. H. (2000b, May). Bayesian mining of spatial lesion-deficit databases. *NeuroImage* **11**(5): S490.
- Herskovits, E. H., V. Megalooikonomou, C. Davatzikos, A. Chen, R. N. Bryan, and J. P. Gerring (1999, November). Is the spatial distribution of brain lesions associated with closed-head injury predictive of subsequent development of attention-deficit/hyperactivity disorder? analysis with brain-image database. *Radiology* **213**(2): 389–394. PMID: 10551217. <http://ditzel.rad.jhu.edu/papers/radiology1999.ps>. ResearchIndex: <http://citeseer.nj.nec.com/319461.html>.
- Hertz, J., A. Krogh, and R. G. Palmer (1991). *Introduction to the Theory of Neural Computation* (1st ed.). Redwood City, California: Addison-Wesley. Santa Fe Institute.
- Hestenes, M. R. and E. Stiefel (1952). Methods of conjugate gradients for solving linear systems. *Journal of Research of the National Bureau of Standards* **49**(6): 409–436.
- Hinton, G. E. (1986). Learning distributed representation of concepts. In *Proceedings of the Eighth Annual conference of the Cognitive Science Society*, Hillsdale, pp. 1–12. Erlbaum.
- Hirsch, M. and S. Koslow (1999, June). Interoperable neuroinformatics management systems for understanding functional neuroanatomy. See Rosen, Seitz, and Volkmann (1999), pp. S70. ISSN 1053–8119.
- Ho, D., D. Feng, and K. Chen (1997, December). Dynamic image data compression in spatial and temporal domains: theory and algorithm. *IEEE Transaction on Information Technology in Biomedicine* **1**(4): 219–228.

- Hoerl, A. E. and R. W. Kennard (1970a, February). Ridge regression: Application to nonorthogonal problems. *Technometrics* **12**(1): 69–82.
- Hoerl, A. E. and R. W. Kennard (1970b, February). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics* **12**(1): 55–67.
- Hofmann, T. (1999). Probabilistic latent semantic analysis. In *Proceedings of the 15th Conference on Uncertainty in AI*, pp. 289–296.
- Hoge, R. D., J. Atkinson, B. Gill, G. R. Crelier, S. Marrett, and G. B. Pike (1999, August). Linear coupling between cerebral blood flow and oxygen consumption in activated human cortex. *Proceedings of the National Academy of Sciences of the USA* **96**(16): 9403–9408. PMID: 10430955. <http://www.pubmedcentral.nih.gov/b.cgi?pubmedid=10430955>.
- Højen-Sørensen, P. A. d. F. R., L. K. Hansen, and C. E. Rasmussen (2000). Bayesian modelling of fMRI time series. Cambridge, Massachusetts, pp. 754–760. MIT Press. <http://eivind.imm.dtu.dk/publications/1999/phs.nips99.ps.gz>.
- Højen-Sørensen, P. A. d. F. R., L. K. Hansen, and E. Rostrup (1999, June). A bayesian approach for estimating activation in fMRI time series. See Rosen, Seitz, and Volkmann (1999), pp. S117. ISSN 1053–8119.
- Holder, D. S. (1987, January). Feasibility of developing a method of imaging neuronal activity in the human brain: a theoretical review. *Med Biol Eng Comput* **25**(1): 2–11.
- Holder, D. S., A. Rao, and Y. Hanquan (1996, November). Imaging of physiologically evoked responses by electrical impedance tomography with cortical electrodes in the anaesthetized rabbit. *Physiol. Meas.* **17**(Supplement 4A): A179–A186. PMID: 9001616.
- Holmes, A. P. and K. J. Friston (1997). Statistical models and experimental design. In *SPMcourse, short course notes*, Chapter 3. Institute of Neurology, Wellcome Department of Cognitive Neurology.
- Holmes, A. P. and K. J. Friston (1998, May). Generalisability, random effects and population inference. See Paus, Gjedde, and Evans (1998). ISSN 1053–8119. <http://www.fil.ion.ucl.ac.uk/spm/RFXabstract.pdf>.
- Holmes, A. P., O. Josephs, C. Büchel, and K. J. Friston (1997, May). Statistical modelling of low-frequency confounds in fMRI. See Friberg, Gjedde, Holm, Lassen, and Nowak (1997), pp. S480. ISSN 1053–8119.
- Hopfinger, J. B., C. Büchel, A. P. Holmes, and K. J. Friston (2000, April). A study of analysis parameters that influence the sensitivity of event-related fmri analyses. *NeuroImage* **11**(4): 326–333.
- Hornak, J. P. (1999). The basics of MRI. <http://www.cis.rit.edu/htbooks/mri>.
- Horwitz, B., K. J. Friston, and J. G. Taylor (2000, October/November). Neural modeling and functional brain imaging: an overview. *Neural Networks* **13**(8–9): 829–846. PMID: 11156195. <http://www.ling.umd.edu/poeppel/HorwitzPET.pdf>.
- Horwitz, B., C. L. Grady, J. V. Haxby, M. B. Schapiro, S. I. Rapoport, L. G. Ungerleider, and M. Mishkin (1992). Functional associations among human posterior extrastriate brain regions during object and spatial vision. *Journal of Cognitive Neuroscience* **4**: 311–322.
- Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology* **24**: 417–441, 498–520.
- Hotelling, H. (1935). The most predicable criterion. *Journal of Educational Psychology* **26**: 139–142.
- Hotelling, H. (1936). Relations between two sets of variates. *Biometrika* **28**: 321–377.
- Hounsfield, G. N. (1973). Computerized transverse axial scanning (tomography): Part 1. description of system. *British Journal of Radiology* **46**: 1016–1022.
- Howard, D., K. Patterson, R. Wise, W. D. Brown, K. J. Friston, C. Weiller, and R. S. J. Frackowiak (1992). The cortical localization of the lexicons. *Brain* **115**: 1769–1782.
- Høy, M., F. Westad, and H. Martens (2001, February). Combining bilinear modeling and ridge regression. Submitted to *Journal of Chemometrics*.
- Hsieh, J.-C., J. Hannerz, and M. Ingvar (1996). Right-lateralised central processing for pain of nitroglycerin-induced cluster headache. *Pain* **67**(1): 59–68. PMID: 8895232.
- Hu, X., T. H. Le, T. Parrish, and P. Erhard (1995, August). Retrospective estimation and correction of physiological fluctuation in functional MRI. *Magnetic Resonance in Medicine* **34**(2): 201–212. PMID: 7476079.

- Huber, J. P. (1972). Robust statistics: A review. *Ann. Math. Statist.* **43**: 1041–1067.
- Hugh, H. M. (2001, July). Method and apparatus for organization and processing information using a digital computer. United States Patent No. 6,256,032.
- Hund-Georgiadis, M., U. Lex, and D. Y. von Cramon (2001, April). Language dominance assessment by means of fMRI: Contributions from task design, performance, and stimulus modality. *Journal of Magnetic Resonance Imaging* **13**(5): 668–675. ISSN 1522-2586. <http://www3.interscience.wiley.com/cgi-bin/abstract/79503512/START>.
- Hurdal, M. K., P. L. Bowers, K. Stephenson, D. L. Sumners, K. Rehm, K. A. Schaper, and D. A. Rottenberg (1999). Quasi-conformally flat mapping the human cerebellum. In C. Taylor and A. Colchester (Eds.), *Medical Image Computing and Computer-Assisted Intervention - MICCAI'99*, Volume 1679 of *Lecture Notes in Computer Science*, Berlin, pp. 279–286. Springer.
- Hurvich, C. M. and C.-L. Tsai (1989). Regression and time series model selection in small samples. *Biometrika* **76**: 297–307.
- Hykin, J., R. Moore, K. Duncan, S. Clare, P. Baker, I. Johnson, R. Bowtell, P. Mansfield, and P. Gowland (1999, August). Fetal brain activity demonstrated by functional magnetic resonance imaging. *Lancet* **354**(9179): 645–646. PMID: 10466668.
- Imaizumi, S., K. Mori, S. Kiritani, R. Kawashima, M. Sugiura, H. Fukuda, K. Itoh, T. Kato, A. Nakamura, K. Hatano, S. Kojima, and K. Nakamura (1997, August). Vocal identification of speaker and emotion activates different brain regions. *Neuroreport* **8**(12): 2809–2812. ISSN 0959-4965. PMID: 9295122.
- Indefrey, P. (2001, June). A meta-analysis of PET and fMRI experiments on syntactic parsing. *NeuroImage* **13**(6): S545. <http://www.apnet.com/www/journal/hbm2001/10136.html>.
- Indefrey, P. and W. J. M. Levelt (1999, June). A meta-analysis of neuroimaging experiments on word production. See Rosen, Seitz, and Volkmann (1999), pp. S1028. ISSN 1053–8119.
- Indefrey, P. and W. J. M. Levelt (2000). The neural correlates of language production. In M. S. Gazzaniga (Ed.), *The New Cognitive Neurosciences* (2nd ed.), Chapter 59, pp. 845–865. Cambridge, MA: MIT Press.
- Ingvar, M., C. Bohm, L. Thurfjell, L. Eriksson, and T. Greitz (1994). The role of a computerized adjustable brain atlas for merging of data from examinations using PET, SPECT, MEG, CT, and MR images. In *Functional neuroimaging*, Chapter 20, pp. 209–215. Academic Press.
- Ingwersen, P. (1998, March). The calculation of web impact factors. *Journal of Documentation* **54**(2): 236–243.
- Institute of Neurology, W. D. o. C. N. (1997, May). <http://www.fil.ion.ucl.ac.uk/spm/course/notes.html>.
- Ionnides, A. A., J. P. Bolton, and C. J. S. Clarke (1990). Continuous probabilistic solutions to the biomagnetic inverse problem. *Inverse Problems*: 523–542.
- Isbell Jr., C. L. and P. Viola (1999). Restructuring sparse high dimensional data for effective retrieval. See Kearns, Solla, and Cohen (1999), pp. 480–486. ISBN 0-262-11245-0. NIPS-11.
- Ishai, A., L. G. Ungerleider, A. Martin, and J. V. Haxby (2000, November). The representation of objects in the human occipital and temporal cortex. *Journal of Cognitive Neuroscience* **12**(suppl. 2): 35–51.
- J. E. Adcock, S. M. S. and P. M. Matthews (2001, June). Comparison of FMRI with the wada test for lateralisation of language in pre-surgical epilepsy patients. *NeuroImage* **13**(6, part 2): S766. <http://www.apnet.com/www/journal/hbm2001/11319.html>.
- Jackson, J. E. (1991). *A user's guide to principal components*. Wiley Serie in Probability and Mathematical Statistics: Applied Probability and Statistics. New York: John Wiley & Sons. ISBN 0-471-62267-2.
- Jenkinson, M. and S. Smith (2001). A global optimisation method for robust affine registration of brain images. *Medical Image Analysis* **5**(2): 143–156.
- Jensen, J. J. (1995, July). 3d visualisering. Technical report, Electronics Institute, Technical University of Denmark, Lyngby, Denmark. <http://eivind.imm.dtu.dk>. In Danish.
- Jensen, M., R. J. Nickles, and S. Holm (1998, May). Carbon-10 dioxide: A new tracer for human brain mapping with PET. See Paus, Gjedde, and Evans (1998), pp. S630. ISSN 1053–8119.

- Jenssen, T.-K., A. Læreid, J. Komorowski, and E. Hovig (2001, May). A literature network of human genes for high-throughput analysis of gene expression. *Nature Genetics* **28**(1): 21–28. PMID: 11326270. http://www.nature.com/cgi-taf/DynaPage.taf?file=/ng/journal/v28/n1/full/ng0501_21.html.
- Jezzard, P. (1999). Basic physiology of fMRI. http://www.fmrib.ox.ac.uk/~peterj/lectures/hbm_2/sld001.htm.
- Jobsis, F. F. (1977). Noninvasive, infrared monitoring of cerebral and myocardial oxygen sufficiency and circulatory parameters. *Science* **198**(4323): 1264–1267.
- Joeri, P., T. Loenneker, T. Huiman, D. Ekatodramis, and E. Martin (1996, June). fMRI of the visual cortex in infants and children. See Belliveau, Fox, Kennedy, Rosen, and Ungeleider (1996). ISSN 1053–8119.
- Johnson, B. and B. Shneiderman (1991, October). Tree-maps: A space-filling approach to the visualization of hierarchical information structures. In *Proceeding of IEEE Visualization '91 Conference, San Diego*, pp. 284–291.
- Johnson, K. A. and J. A. Becker (1997, December). The whole brain atlas v1.0. CDROM/Internet. <http://www.med.harvard.edu/AANLIB/home.html>.
- Johnsrude, I., R. Cusack, P. Morosan, D. Halland, M. Brett, K. Zilles, and R. S. J. Frackowiak (2001, June). Cytoarchitectonic region-of-interest analysis of auditory imaging data. *NeuroImage* **13**(6): S897. <http://www.apnet.com/www/journal/hbm2001/10458.html>.
- Jones, R. A. (1999, August). Origin of the signal undershoot in BOLD studies of the visual cortex. *NMR in Biomedicine* **12**(5): 301–308. PMID: 10484819. <http://www3.interscience.wiley.com/cgi-bin/abstract/64500046/START>.
- Jones, T., D. A. Chesler, and M. M. Ter-Pogossian (1976, April). The continuous inhalation of oxygen-15 for assessing regional oxygen extraction in the brain of man. *Br. J. Radiol.* **49**(580): 339–343. PMID: 820399.
- Jöreskog, K. G. (1967). Some contributions to maximum likelihood factor analysis. *Psychometrika* **32**: 443–482.
- Jöreskog, K. G. (1970). A general methods for analysis of covariance structures. *Biometrika* **57**: 239–251.
- Josephs, O. and R. N. A. Henson (1999, July). Event-related functional magnetic resonance imaging: modelling, inference and optimization. *Philosophical Transaction of the Royal Society of London, Series B* **354**(1387): 1228–1228. PMID: 10466147.
- Josephs, O., R. Turner, and K. J. Friston (1997). Event-related fMRI. *Human Brain Mapping* **5**(4): 243–248.
- Jueptner, M., M. Rijntjes, C. Weiller, J. H. Faiss, D. T. S. P. Mueller, and H. C. Diener (1995, August). Localization of a cerebellar timing process using PET. *Neurology* **45**(8): 1540–1545. PMID: 7644055.
- Kabani, N., G. L. Goualher, D. MacDonald, and A. C. Evans (2001, February). Measurement of cortical thickness using an automated 3-D algorithm: A validation study. *NeuroImage* **13**(2): 375–380.
- Karhunen, K. (1947). Über lineare metoden in der wahrscheinlichkeitsrechnung. *Ann. Acad. Sci Fennicae, Ser. A137* **37**: 3–79.
- Kato, H., T. Nakayama, and Y. Yamane (2000, August). Navigation analysis tool based on the correlation between contents distribution and access patterns. In *Workshop on Web Mining for E-Commerce — Challenges and Opportunities (WEBKDD'2000)*. ACM. <http://robotics.stanford.edu/~Eronnyk/WEBKDD2000/papers/kato.pdf>. ResearchIndex: <http://citeseer.nj.nec.com/354234.html>.
- Kato, T., A. Kamei, S. Takashima, and T. Ozaki (1993, May). Human visual cortical function during photic stimulation monitoring by means of near-infrared spectroscopy. *Journal of Cerebral Blood Flow and Metabolism* **13**(3): 516–520. PMID: 8478409.
- Kato, T., J. Ohyu, M. Fukumizu, and S. Takashima (1998, May). Assessment of language lateralization using functional near-infrared spectroscopy in bedside. See Paus, Gjedde, and Evans (1998), pp. S207. ISSN 1053–8119.
- Kearns, M. S., S. A. Solla, and D. Cohen (Eds.) (1999). Cambridge, Massachusetts. MIT Press. ISBN 0-262-11245-0. NIPS-11.
- Kelle, O. (1999). Remote brain image segmentation. In *Conference Proceedings IT IS-ITAB '99. Joint Meeting. Second International Workshop On the Telemedical Information Society (IT IS '99)/Second IEEE EMBS International Workshop on Information Technology Applications in Biomedicine (ITAB '99) (Cat. No.99TH8457)*, pp. 22. IEEE. ISBN 0780356470.

- Kendall, D. G. (1971). Seriation from abundance matrices. In F. R. Hodson, D. G. Kendall, and P. Tantu (Eds.), *Mathematics in the Archeological and Historical Sciences*, pp. 215–251. Edinburgh University Press.
- Kennan, R. P. and R. T. Constable (2001, June). Comparison of near infrared optical encephalography and fMRI for assessment of language dominance. *NeuroImage* **13**(6, part 2): S22. <http://www.apnet.com/www/journal/hbm2001/11807.html>.
- Kennedy, P. R., R. A. E. Bakay, M. M. Moore, K. Adams, and J. Goldwaithe (2000, June). Direct control of a computer from the human central nervous system. *IEEE Transactions on Rehabilitation Engineering*. **8**(2): 198–202. PMID: 10896186.
- Kepler, J. (1609). *Astronomia nova*. Ulm, Germany.
- Kepler, J. (1619). *Harmonices Mundi*. Linz, Germany.
- Kepler, J. (1627). *Rudolphine Tables*.
- Kershaw, J., B. A. Ardekani, and I. Kanno (1999, December). Application of Bayesian inference to fMRI data analysis. *IEEE Transaction on Medical Imaging* **18**(12). PMID: 10695527.
- Keskin, C. and V. Vogelmann (1997, November). Effective Visualization of Hierarchical Graphs with the Cityscape Metaphor. See Ebert and Nicholas (1997). ISBN 1581130511.
- Kessler, M. M. (1963, January). Bibliographic coupling between scientific papers. *American Documentation* **14**(1): 10–25.
- Kettenring, J. R. (1971). Canonical analysis of several sets of variables. *Biometrika* **58**: 433–451.
- Kety, S. S. and C. F. Schmidt (1945). The determination of cerebral blood flow in man by the use of nitrous oxide in low concentrations. *Am. J. Physiol.* **143**: 53–66.
- Kevles, B. H. (1998). *Naked to the bone: Medical imaging in the twentieth century*. Sloan Technology Series. Reading, Massachusetts: Addison-Wesley. ISBN 0-201-32833X.
- Kiebel, S. J., J. Ashburner, J.-B. Poline, and K. J. Friston (1997, May). MRI and PET coregistration — a cross validation of statistical parametric mapping and automated image registration. *NeuroImage* **5**(4): 271–279. PMID: 9345556.
- Kiers, H. A. L. (2000). Towards a standardized notation and terminology in multiway analysis. *Journal of Chemometrics* **14**(3): 105–122. <http://www3.interscience.wiley.com/cgi-bin/abstract/72504977/START>, <http://www.spectroscopynow.com/Spy/pdfs/kiers.pdf>.
- Kinahan, P. E. and D. C. Noll (1999, April). A direct comparison between whole-brain PET and BOLD fMRI measurements of single subject activation response. *NeuroImage* **9**(4): 430–438. PMID: 10191171.
- Kingston, J. H. (1990). *Algorithms and data structures: Design, correctness, analysis*. Addison-Wesley.
- Kirkland, K. (2001). Analysis tools of mulab. Internet. <http://mulab.physiol.upenn.edu/analysis.html>.
- Kjell, B. and O. Frieder (1992, October). Visualization of literary style. In *IEEE International Conference on Systems, Man and Cybernetics*.
- Kjems, U., L. K. Hansen, S. Muley, J. R. Anderson, D. A. Rottenberg, S. A. Frutiger, and S. C. Strother (2000, May). Learning curves for evaluation of neuroimaging experiments. See Fox and Lancaster (2000), pp. S637. ISSN 1053-8119. <http://www.apnet.com/www/journal/hbm2000/6332.html>.
- Kjems, U., S. C. Strother, J. R. Anderson, and L. K. Hansen (1999, April). Enhancing the multivariate signal of [15O] water PET studies with a new nonlinear neuroanatomical registration algorithm. *IEEE Transaction on Medical Imaging* **18**(4): 306–319. PMID: 10385288.
- Kleinberg, J. M. (1997, May). Authoritative sources in a hyperlinked environment. Technical Report 10076(91892), IBM. ResearchIndex: <http://citeseer.nj.nec.com/kleinberg97authoritative.html>.
- Kleinberg, J. M. (1998). Authoritative sources in a hyperlinked environment. In *Proceedings of the Ninth Annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 668–677.
- Kleinberg, J. M. (1999). Authoritative sources in a hyperlinked environment. *Journal of ACM* **46**(5): 604–632.
- Kling-Petersen, T., R. Pascher, and M. Rydmark (1998). 3d-brain 2.0—narrowing the gap between personal computers and high end workstations. *Stud. Health. Technol. Inform.* **50**: 234–239. PMID: 10180546. <http://www.mice.gu.se/MICE/pubs/MMVR6tkp.pdf>.

- König, A. H., H. Doleisch, and E. Gröller (1999, February). Multiple views and magic mirrors — fMRI visualization of the human brain. Technical Report TR-186-2-99-08, Institute of Computer Graphics and Algorithms, Vienna University of Technology, A-1040 Karlsplatz 13/186/2. <ftp://ftp.cg.tuwien.ac.at/pub/TR/99/TR-186-2-99-08Paper.ps.gz>. ResearchIndex: <http://citeseer.nj.nec.com/35545.html>.
- Knight, J. R. (1995). Super-pattern matching. *Algorithmica* **13**(1/2): 211–243.
- Knowles, A. J., D. J. Manton, B. Issa, and L. W. Turnbull (1998a, April). Application of neural networks to the analysis of fMRI time-course data reveals increased sensitivity to functional activation. In *International Society for Magnetic Resonance in Medicine Sixth Annual Meeting*, pp. 1477.
- Knowles, A. J., D. J. Manton, and L. W. Turnbull (1998b, December). Application of neural networks to fMRI: increased sensitivity at low CNR compared to statistical methods. In *Fourth Annual General Meeting of the British Chapter of ISMRM*.
- Kochunov, P., J. L. Lancaster, and P. T. Fox (2000, May). Java based platform independent visual interface for reporting brain-mapping studies. See Fox and Lancaster (2000). ISSN 1053-8119.
- Kohonen, T. (1995). *Self-Organizing Maps*. Berlin: Springer-Verlag.
- Kolenda, T. (1998, September). Independent component analysis. Master's thesis, Department of Mathematical Modelling, Technical University of Denmark, Lyngby, Denmark. IMM-EKS-1998-39, In Danish.
- Kolenda, T. and L. K. Hansen (1999, May). Independent components in text. Submitted to NIPS'99, but not accepted.
- Kollokian, V. (1996, November). Performance analysis of automatic techniques for tissue classification in MRI of the human brain. Master's thesis, Concordia University, Montreal, Canada.
- Konishi, S., D. I. Donaldson, and R. L. Buckner (2001, February). Transient activation during block transition. *NeuroImage* **13**(2): 364–374.
- Koski, L. and T. Paus (2000, July). Functional connectivity of the anterior cingulate cortex within the human frontal lobe: a brain-mapping meta-analysis. *Experimental Brain Research* **133**(1): 55–65. PMID: 10933210.
- Koslow, S. H. and M. F. Huerta (Eds.) (1997, February). *Neuroinformatics, An Overview of the Human Brain Project*. A Volume in the Progress in Neuroinformatics Research Series. Lawrence Erlbaum Associates. ISBN 080582099X.
- Kosslyn, S. M., N. M. Alpert, W. L. Thompson, C. F. Chabris, S. L. Rauch, and A. K. Anderson (1994). Identifying objects seen from different viewpoints, a PET investigation. *Brain* **117**: 1055–1071.
- Kosslyn, S. M., W. T. Thompson, and N. M. Albert (1995). Identifying objects at different levels of hierarchy: A positron emission tomography study. *Human Brain Mapping* **3**: 107–132.
- Kotz, S., N. L. Johnson, and C. B. Read (Eds.) (1982). *Encyclopedia of Statistical Science*. John Wiley & Sons. ISBN 0-471-05546-8.
- Koutsofios, E. and S. C. North (1996, November). *Drawing graphs with dot*. Murray Hill, New Jersey: AT&T Bell Laboratories.
- Kruger, A., C. L. Giles, F. M. Coetzee, E. Glover, G. W. Flake, S. Lawrence, and C. Omlin (2000, November). DEADLINER: Building a new niche search engine. In *Ninth International Conference on Information and Knowledge Management, CIKM 2000, Washington, DC*. ResearchIndex: <http://citeseer.nj.nec.com/kruger00deadliner.html>.
- Krüger, G., A. Kleinschmidt, and J. Frahm (1996, Jun). Dynamic MRI sensitized to cerebral blood oxygenation and flow during sustained activation of human visual cortex. *Magnetic Resonance in Medicine* **35**(6): 797–800. PMID: 8744004.
- Kruggel, F., X. Decombes, and D. Y. von Cramon (1998). Preprocessing of fMR datasets. In B. Vemuri (Ed.), *Proceedings. Workshop on Biomedical Image Analysis*, pp. 211–220. IEEE Computer Society.
- Kruggel, F. and D. Y. von Cramon (1999, October). Modeling the hemodynamic response in single-trial functional MRI experiments. *Magnetic Resonance in Medicine* **42**(4): 787–797. PMID: 10502769. <http://www3.interscience.wiley.com/cgi-bin/abstract/66000379/START>.
- Kruggel, F., S. Zysset, and D. Y. von Cramon (2000, August). Nonlinear regression of functional MRI data: An item recognition task study. *NeuroImage* **12**(2): 173–183.
- Krutchkoff, R. G. (1967, August). Classical and inverse regression methods of calibration. *Technometrics* **9**(3): 425–439.

- Krzanowski, W. J. (1979, September). Between-groups comparison of principal components. *Journal of the American Statistical Association* **74**(367): 703–707.
- Krzanowski, W. J., P. Jonathan, W. V. McCarthy, and M. R. Thomas (1995). Discriminant analysis with singular covariance matrices: Methods and applications to spectroscopic data. *Applied Statistics* **44**(1): 101–115.
- Kübler, A., B. Kotchoubey, T. Hinterberger, N. Ghanayim, J. Perelmouter, M. Schauer, C. Fritsch, E. Taub, and N. Birbaumer (1999, January). The thought translation device: a neurophysiological approach to communication in total motor paralysis. *Experimental Brain Research* **124**(2): 223–232.
- Kulick, T. (1996). Building an OpenGL volume renderer. Technical report, Silicon Graphics Inc.
- Kullback, S. and R. A. Leibler (1951). On information and sufficiency. *Annals of Mathematical Statistics* **22**: 79–86.
- Kung, S. Y. and K. I. Diamantaras (1990). A neural network learning algorithm for adaptive principal component extraction (APEX). In *International Conference on Acoustics, Speech, and Signal Processing*, pp. 861–864. Albuquerque, NM.
- Kuroda, K., K. Oshio, T. Sakamoto, N. Takei, T. Nakai, T. Okada, K. Yanaka, and A. Matsumura (2001, June). Internally-referenced temperature imaging using line scan echo planar spectroscopic imaging. *NeuroImage* **13**(6): S23.
- Kustra, R. (2000, August). *Statistical Analysis of Medical Images with Applications to Neuroimaging*. Ph. D. thesis, Graduate Department of Public Health Sciences, University of Toronto, Toronto, Ontario. <http://www.utstat.toronto.edu/~rafal/thesis.ps.gz>.
- Kustra, R. and S. C. Strother (2000, December). Penalized discriminant analysis of [¹⁵O]-water PET brain images with prediction error selection of smoothness and regularization hyperparameters. Preprint.
- Kwan, R. K.-S., A. C. Evans, and G. B. Pike (1996, December). An extensible MRI simulator for post-processing evaluation. In K. H. Höhne and R. Kikinis (Eds.), *Visualization in Biomedical Computing, 4th International Conference, VBC '96, Hamburg, Germany, September 22–25, 1996, Proceedings*, Volume 1131 of *Lecture Notes in Computer Science*, Heidelberg, Germany, pp. 135–140. Springer Verlag. ISBN 3540616497. <http://www.bic.mni.mcgill.ca/users/rkwan/vbc96/paper/vbc96.ps.gz>.
- Kwan, R. K.-S., A. C. Evans, and G. B. Pike (1999, November). MRI simulation-based evaluation of image-processing and classification methods. *IEEE Transactions on Medical Imaging* **18**(11): 1085–1097. PMID: 10661326.
- Kwong, K. K., J. W. Belliveau, D. A. Chesler, I. E. Goldberg, R. M. Weisskoff, B. P. Poncelet, D. N. Kennedy, B. E. Hoppel, M. S. Cohen, R. Turner, H.-M. Cheng, T. J. Brady, and B. R. Rosen (1992). Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation. *Proceedings of the National Academy of Sciences* **89**: 5675–5679.
- Lai, P. L. and C. Fyfe (1999). A neural implementation of canonical correlation analysis. *Neural Networks* **12**(10): 1391–1397. <http://cis.paisley.ac.uk/fyfe-ci0/cca/nn.zip>.
- Lamping, J. and R. Rao (1994, November). Laying out and visualizing large trees using a hyperbolic space. pp. 13–14. ACM Press.
- Lamping, J. and R. Rao (1996). The hyperbolic browser: A focus + context technique for visualizing large hierarchies. *Journal of Visual Languages and Computing* **7**(1): 33–35.
- Lamping, J., R. Rao, and P. Pirolli (1995, May). A focus+context technique based on hyperbolic geometry for visualizing large hierarchies. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems, Denver, Colorado*. ACM. http://www.acm.org/sigchi/chi95/proceedings/papers/jl_bdy.htm.
- Lamping, J. O. and R. B. Rao (1997, April). Displaying node link structure for organisation charts, file system hierarchies, hyper-text, world wide web and SGML structures - using inner convex hulls of first and last representations which include subsets of node features representing different sets of nodes. United States Patent No. US5619632, EP702331-A2, JP08110847-A.
- Lancaster, J. L., E. Chan, S. A. Mikiten, S. Nguyen, and P. T. Fox (1997a, May). BrainMap™ search and view. See Friberg, Gjedde, Holm, Lassen, and Nowak (1997), pp. S634. ISSN 1053–8119.
- Lancaster, J. L., P. T. Fox, G. Davis, and S. Mikiten (1994, February). BrainMap: A database of human functional brain mapping. In *The Fifth International Conference: Peace through Mind/Brain Science*, Hammamatsu, Japan. <http://ric.uthscsa.edu>.

- Lancaster, J. L., T. G. Glass, B. R. Lankipalli, H. Downs, H. Mayberg, and P. T. Fox (1995). A modality-independent approach to spatial normalization of tomographic images of the human brain. *Human Brain Mapping* **3**: 209–223.
- Lancaster, J. L., P. Kochunov, and P. T. Fox (2001, June). An individual representative target brain in Talairach space. *NeuroImage* **13**(6, part 2): S180. <http://www.apnet.com/www/journal/hbm2001/11325.html>.
- Lancaster, J. L., P. Kochunov, D. Nickerson, and P. T. Fox (2000a, May). Stand-alone java-base version of the Talairach Daemon database system. See Fox and Lancaster (2000), pp. S923. ISSN 1053-8119.
- Lancaster, J. L., P. Kochunov, M. Woldorff, M. Liotti, M. Parsons, L. Rainey, D. Nickerson, and P. T. Fox (2000b, May). Automatic Talairach labels for functional activation sites. See Fox and Lancaster (2000), pp. S483. ISSN 1053-8119.
- Lancaster, J. L., L. H. Rainey, J. L. Summerlin, C. S. Freitas, P. T. Fox, A. C. Evans, A. W. Toga, and J. C. Mazziotta (1997b). Automated labeling of the human brain: A preliminary report on the development and evaluation of a forward-transform method. *Human Brain Mapping* **5**(4): 238–242. ISSN 1065-9471. <http://www3.interscience.wiley.com/cgi-bin/abstract/56443/START>.
- Lancaster, J. L., J. L. Summerlin, L. Rainey, C. S. Freitas, and P. T. Fox (1997c, May). The Talairach daemon, a database server for Talairach atlas labels. See Friberg, Gjedde, Holm, Lassen, and Nowak (1997), pp. S633. ISSN 1053-8119.
- Lancaster, J. L., M. G. Woldorff, M. Liotti, C. S. Freitas, L. Rainey, P. V. Kochunov, D. Nickerson, S. A. Mikiten, and P. T. Fox (2000c, July). Automated Talairach atlas labels for functional brain mapping. *Human Brain Mapping* **10**(3): 120–131. PMID: 10912591.
- Lange, N. (1996). Tutorial on biostatistics: statistical approaches to human brain mapping by functional magnetic resonance imaging. *Statistics in Medicine* **15**: 389–428.
- Lange, N., L. K. Hansen, J. R. Anderson, F. Å. Nielsen, R. Savoy, S.-G. Kim, and S. C. Strother (1998, May). An empirical study of statistical model complexity in neuro-fMRI. See Paus, Gjedde, and Evans (1998). ISSN 1053-8119.
- Lange, N., L. K. Hansen, M. W. Pedersen, R. L. Savoy, and S. C. Strother (1996, June). A concordance correlation coefficient reproducibility of spatial activation patterns. See Belliveau, Fox, Kennedy, Rosen, and Ungeleider (1996), pp. S75. ISSN 1053-8119.
- Lange, N., S. C. Strother, J. R. Anderson, F. Å. Nielsen, A. P. Holmes, T. Kolenda, R. Savoy, and L. K. Hansen (1999, September). Plurality and resemblance in fMRI data analysis. *NeuroImage* **10**(3): 282–303. PMID: 10458943. <http://www.idealibrary.com/links/citation/1053-8119/10/282>.
- Lange, N. and S. L. Zeger (1997). Non-linear fourier time series analysis for human brain mapping by functional magnetic resonance imaging (with discussion). *Applied Statistics, Journal of the Royal Statistical Society, Series C* **46**(1): 1–29.
- Laplane, D., J. Talairach, V. Meininger, J. Bancaud, and A. Bouchareine (1977, January-February). Motor consequences of motor area ablations in man. *J. Neurol. Sci.* **31**(1): 29–49. PMID: 833609.
- Larkey, L. S., P. Ogilvie, M. A. Price, and B. Tamillo (2000). Acrophile: An automated acronym extractor and server. In *Proceedings of the 5th ACM International Conference on Digital Libraries (DL'00)*.
- Larsen, J. (1996, January). *Design of Neural Network Filters*. Ph. D. thesis, Electronics Institute, Technical University of Denmark, Lyngby, Denmark. ftp://eivind.imm.dtu.dk/dist/PhD_thesis/jlarsen.thesis.pdf. Second edition.
- Larsen, J., L. N. Andersen, M. Hintz-Madsen, and L. K. Hansen (1998a). Design of robust neural network classifiers. In *Proceedings of ICASSP'98*, Volume 2, pp. 1205–1208. Seattle, USA. <http://eivind.imm.dtu.dk/publications/1998/larsen.icassp98.ps.Z>.
- Larsen, J. and L. K. Hansen (1994). Generalization performance of regularized neural network models. See Vrontzos, Hwang, and Wilson (1994), pp. 42–51. <http://eivind.imm.dtu.dk>.
- Larsen, J. and L. K. Hansen (1995). Empirical generalization assessment of neural network models. In F. Girosi (Ed.), *Neural Networks for Signal Processing V*, pp. 42–51. IEEE: IEEE Press. ISBN 078032739X. <http://eivind.imm.dtu.dk/publications/1995/larsen.nnsp95.ps.Z>.
- Larsen, J., C. Svarer, L. N. Andersen, and L. K. Hansen (1998b). Adaptive regularization in neural network modeling. In K. M. G.B. Orr (Ed.), *Neural Networks: Tricks of the Trade*, Volume 1524 of *Lecture Notes in Computer Science*. Germany: Springer-Verlag. ISBN 3-540-65311-2. <http://eivind.imm.dtu.dk/publications/1998/larsen.bot.ps.Z>.

- Larson, R. R. (1996). Bibliometrics of the world wide web: An exploratory analysis of the intellectual structure of cyberspace. In *Proceedings of the 1996 Annual ASIS Meeting, Baltimore*, pp. 71–79. <http://sherlock.berkeley.edu/asis96/asis96.html>.
- Larsson, S. and P. Hansson (2000). Information views: Support in the design process. In L. Svensson, U. Snis, C. Sørensen, H. Fägerlind, T. Lindroth, M. Magnusson, and C. Östlund (Eds.), *Proceedings of IRIS 23*. Laboratorium for Interaction Technology, University of Trollhättan Uddevalla. <http://iris23.htu.se/proceedings/PDF/99final.PDF>. ResearchIndex: <http://citeseer.nj.nec.com/420061.html>.
- Lassen, N. A., D. H. Ingvar, and E. Skinhøj (1978, October). Brain function and blood flow. *Scientific American* **239**(4): 62–71. PMID: 705327.
- Lauterbur, P. G. (1973). Image formation by induced local interactions: Examples employing nuclear magnetic resonance. *Nature* **242**: 190–191.
- Lautrup, B., L. K. Hansen, I. Law, N. J. S. Mørch, C. Svarer, and S. C. Strother (1995). Massive weight sharing: A cure for extremely ill-posed problems. In H. J. Herman, D. E. Wolf, and E. P. Poppel (Eds.), *Proceedings of the Workshop on Supercomputing in Brain Research: From Tomography to Neural Networks*, Singapore, pp. 137–148. World Scientific. <http://eivind.imm.dtu.dk/publications/1994/lautrup.massive.ps.Z>.
- Law, I. (1996). *Saccadic Eye Movements, A Functional Brain Mapping Approach*. Ph. D. thesis, Department of Neurology, Rigshospitalet, The University of Copenhagen, Copenhagen, Denmark. <http://nru.dk/cgi-bin/extract?lawi1997c>.
- Law, I., I. Kanno, H. Fuhita, S. Miura, N. A. Lassen, and K. Uemura (1991). Left supramarginal/angular gyri activation during reading of syllabograms in the Japanese language. *Journal Neurolinguistics* **6**(3): 243–251.
- Law, I., C. Svarer, S. Holm, and O. B. Paulson (1997a, November). The activation pattern in normal man during suppression, imagination and performance of saccadic eye movements. *Acta Physiol Scand* **161**(3): 419–434.
- Law, I., C. Svarer, and O. B. Paulson (1996). Occipito-parietal cortex activation during self-generated eye movements. In *Second Congress of the European Federation of Neurological Sciences 1996*, Volume 3 suppl. 5, pp. P169.
- Law, I., C. Svarer, and O. B. Paulson (1997b, May). Sensitivity to movement rate and exposure duration of oculomotor system activation. See Friberg, Gjedde, Holm, Lassen, and Nowak (1997), pp. S250. ISSN 1053–8119.
- Law, I., C. Svarer, E. Rostrup, and O. B. Paulson (1998, November). Parieto-occipital cortex activation during self-generated eye movements in the dark. *Brain* **121**(11): 2189–2200. PMID: 9827777.
- Lawrence, S., F. Coetzee, G. W. Flake, D. M. Pennock, R. Krovetz, F. Å. Nielsen, A. Kruger, and C. L. Giles (2000). Persistence of information on the web: Analyzing citations contained in research article. In *Proceedings of the 2000 ACM CIKM International Conference on Information and Knowledge Management, McLean, VA, USA, November 6–11, 2000*, pp. 235–242. ACM. <http://www.acm.org/pubs/citations/proceedings/cikm/354756/p235-lawrence/>.
- Lawrence, S., F. Coetzee, E. Glover, D. M. Pennock, G. Flake, F. Å. Nielsen, R. Krovetz, A. Kruger, and C. L. Giles (2001). Persistence of web references in scientific research. *IEEE Computer* **34**(2): 26–31. <http://www.neci.nec.com/~lawrence/papers/persistence-computer01/persistence-computer01.ps.gz>.
- Lawrence, S. and C. L. Giles (1998a). Context and page analysis for improved web search. *IEEE Internet Computing* **2**(4): 38–46.
- Lawrence, S. and C. L. Giles (1998b, April). Searching the world wide web. *Science* **280**(5360): 98–100. PMID: 9525866.
- Lawrence, S., C. L. Giles, and K. Bollacker (1999). Digital libraries and autonomous citation indexing. *IEEE Computer* **32**(6): 67–71.
- Le Cun, Y., J. S. Denker, and S. A. Solla (1990). Optimal brain damage. In D. S. Touretzky (Ed.), *Advances in Neural Information Processing Systems: Proceedings of the 1989 Conference*, San Mateo, CA, pp. 598–605. Morgan-Kaufmann. NIPS-2.
- Leach, M. O. (1988). Spatially localised nuclear. In S. Webb (Ed.), *The Physics of medical Imaging*, Chapter 8, pp. 389–487. Bristol: IOP Publishing. ISBN 0-85274-349.
- Leblanc, R., E. Meyer, D. Bub, R. J. Zatorre, and A. C. Evans (1992, August). Language localization with activation positron emission tomography scanning. *Neurosurgery* **31**(2): 369–373. PMID: 1513446.

- Ledberg, A. (1998, May). Comparing patterns of functional covariations between several groups. See Paus, Gjedde, and Evans (1998), pp. S752. ISSN 1053-8119.
- Ledberg, A. (2000, March). Robust estimation of the probabilities of 3-d clusters in functional brain images: application to pet data. *Human Brain Mapping* **9**(3): 143-155. PMID: 10739365.
- Ledberg, A., S. Åkerman, and P. E. Roland (1998, August). Estimation of the probability of 3D clusters in functional brain images. *NeuroImage* **8**(2): 111-126. <http://www.idealibrary.com/links/doi/10.1006/nimg.1998.0336>.
- Lee, D. D. and H. S. Seung (1999, October). Learning the parts of objects by non-negative matrix factorization. *Nature* **401**: 788-791.
- Lee, D. D. and H. S. Seung (2001). Algorithms for non-negative matrix factorization.
- Lee, T.-W. (1998). *Independent Component Analysis, Theory and Applications*. Kluwer Academic Publishers.
- Leen, T. K., T. G. Dietterich, and V. Tresp (Eds.) (2001). *Advances in Neural Information Processing Systems*, Boston, MA. MIT Press. NIPS-13.
- Leibovici, D. G. (2000). Multiway multidimensional analysis for pharmaco-EEG studies. Submitted. <http://www.fmrib.ox.ac.uk/~didier/cv/pap/multteeg.ps>.
- Lepage, M., O. Ghaffar, L. Nyberg, and E. Tulving (2000, January). Prefrontal cortex and episodic memory retrieval mode. *Proceedings of the National Academy of Sciences of the USA* **97**(1): 506-511. PMID: 10618448. <http://www.pnas.org/cgi/content/full/97/1/506>.
- Lepage, M., R. Habib, and E. Tulving (1998). Hippocampal PET activations of memory encoding and retrieval: The HIPER model. *Hippocampus* **8**(4): 313-322.
- Letovsky, S. I., S. H. Whitehead, C. H. Paik, G. A. Miller, J. Gerber, E. H. Herskovits, T. K. Fulton, and R. N. Bryan (1998, November-December). A brain image database for structure/function analysis. *AJNR American Journal of Neuroradiology* **19**(10): 1869-1877. PMID: 9874539. <http://www.ajnr.org/cgi/reprint/19/10/1869.pdf>.
- Levenberg, K. (1944). A method for the solution of certain non-linear problems in least squares. *Quarterly Journal of Applied Mathematics* **II**(2): 164-168.
- Levin, A. U., T. K. Leen, and J. E. Moody (1994). Fast pruning using principal components. See Cowan, Tesauero, and Alspector (1994), pp. 35-42. NIPS-6.
- Liao, C. H., K. J. Worsley, J.-B. Poline, G. H. Duncan, and A. C. Evans (2001a, June). Estimating the delay of the hemodynamic response in fMRI data. *NeuroImage* **13**(6, part 2): S185. <http://www.apnet.com/www/journal/hbm2001/10799.html>.
- Liao, C. H., K. J. Worsley, J.-B. Poline, G. H. Duncan, and A. C. Evans (2001b, May). Estimating the delay of the response in fMRI data. Preprint. <http://euclid.math.mcgill.ca/keith/delay/delay.pdf>. Submitted to NeuroImage.
- Liao, Y.-K. (1998). Effects of hypermedia versus traditional instruction on students' achievement: a meta-analysis. *Journal of Research on Computing in Education* **30**(4): 341-59.
- Lin, J. C., J. R. Gates, F. J. Ritter, M. B. Dunner, X. Hu, D. E. Miulli, D. D. Roman, C. A. Nelson, and C. L. Truwit (1997, May). Functional mapping of Wernicke's and Broca's area during language processing. See Friberg, Gjedde, Holm, Lassen, and Nowak (1997), pp. S583. ISSN 1053-8119.
- Lindley, D. V. and A. F. M. Smith (1972). Bayes estimates for the linear model (with discussion). *Journal Royal Statistical Society, Series B* **34**: 1-41.
- Lipschutz, B., J. Ashburner, K. J. Friston, and C. J. Price (2000, May). Assessing study-specific regional variation in fMRI signal. See Fox and Lancaster (2000), pp. S460. ISSN 1053-8119.
- Lipshutz, B., K. J. Friston, J. Ashburner, and C. J. Price (2001, February). Assessing study-specific regional variations in fMRI signal. *NeuroImage* **13**(2): 392-398.
- Liu, Y. and F. Dellaert (1998a, June). A classification based similarity metric for 3d image retrieval. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'98)*, Santa Babara, CA, pp. 800-805. <http://www.cs.cmu.edu/afs/cs.cmu.edu/project/nist/ftp/cvpr98.ps.Z>.
- Liu, Y. and F. Dellaert (1998b, November). Classification-driven medical image retrieval. In *DAPAR Image Understanding Workshop, IUW'98*. <http://www.cs.cmu.edu/afs/cs.cmu.edu/project/nist/ftp/iuw.ps.Z>.

- Lloyd, D. (1999, June). Terra cognita: From functional neuroimaging to the map of the mind. <http://www.trincoll.edu/~dlloyd/terra.html>. (Submitted to Brain and Mind).
- Lloyd, D. (2000, May). Multivariate meta-analysis of studies in the brainmap archive. See Fox and Lancaster (2000), pp. S911. ISSN 1053-8119.
- Lobaugh, N. J., R. West, and A. R. McIntosh (2001, May). Spatiotemporal analysis of experimental differences in event-related potential data with partial least squares. *Psychophysiology* **38**(3): 517–530. PMID: 11352141.
- Loève, M. (1963). *Probability Theory*. New York: Van Nostrand.
- Loh, W.-L. (1995). On linear discriminant analysis with adaptive ridge classification rules. *Journal of Multivariate Analysis* **53**: 264–278.
- Lohmann, G., K. Müller, V. Bosch, H. Mentzel, N. Busch, and D. Y. von Cramon (2001, June). Lipsia - a software package for the analysis of fMRI data. *NeuroImage* **13**(6): S190. <http://www.apnet.com/www/journal/hbm2001/10174.html>.
- Longstaff, A. (2000). *Neuroscience*. Instant Notes. Oxford, UK: BIOS Scientific Publishers Limited. ISBN 1-85996-082-0.
- Longstreth Jr, W. T., P. Diehr, T. A. Manolio, N. J. Beauchamp, C. A. Jungreis, and D. Lefkowitz (2001). Cluster analysis and patterns of findings on cranial magnetic resonance imaging of the elderly: the cardiovascular health study. *Arch Neurol*. **58**(4): 635–640. PMID: 11295995.
- Lorensen, W. E. and H. E. Cline (1987, July). Marching cubes: A high resolution 3D surface construction algorithm. *Computer graphics* **21**(4): 163–168.
- Lowe, M. J., M. Dziedzic, J. T. Lurito, V. P. Matthews, and M. D. Philips (2000, May). Temporal correlations in low frequency BOLD fluctuations reveal functional networks. See Fox and Lancaster (2000), pp. S510. ISSN 1053-8119. <http://www.academicpress.com/www/journal/hbm2000/6331.html>.
- Lowe, M. J. and J. A. Sorensen (1997, May). Spatially filtering functional magnetic resonance imaging data. *Magnetic Resonance in Medicine* **37**(5): 723–729.
- Lowry, O. H., N. J. Rosebrough, A. L. Farr, and R. J. Randall (1951). Protein measurement with folin phenol reagent. *Journal of Biological Chemistry* **193**: 265–275.
- Luhn, H. P. (1957, October). A statistical approach to the mechanized encoding and searching of literary information. *IBM Journal of Research and Development* **1**(4): 309–317.
- Lund, T. E. and L. G. Hanson (2001, June). Physiological noise reduction in fMRI using vessel time-series as covariates in a general linear model. *NeuroImage* **13**(6, Part 2): S191. <http://www.apnet.com/www/journal/hbm2001/10608.html>.
- Lundsager, B. and B. L. Kristensen (1996, Feb). Lineær og ulineær modellering af positron emission tomografier. Master's thesis, Department of Mathematical Modelling, Technical University of Denmark, Lyngby, Denmark. <http://eivind.imm.dtu.dk>. In Danish.
- Lunin, L. F. and H. D. White (1990). Author cocitation analysis, introduction. *Journal of the American Society for Information Science* **41**(6): 429–432.
- MacDonald, D., N. Kabani, D. Avis, and A. C. Evans (2000, September). Automated 3-D extraction of inner and outer surfaces of cerebral cortex from MRI. *NeuroImage* **12**(3): 340–356. PMID: 10944416.
- MacKay, D. J. C. (1992a). Bayesian interpolation. *Neural Computation* **4**(3): 415–447.
- MacKay, D. J. C. (1992b). Information-based objective functions for active data selection. *Neural Computation* **4**(4): 590–604. <ftp://wol.ra.phy.cam.ac.uk/pub/www/mackay/selection.nc.ps.gz>. ResearchIndex: <http://citeseer.nj.nec.com/47461.html>.
- MacKay, D. J. C. (1994). Bayesian methods for backpropagation networks. In E. Dormann, J. L. van Hemmen, and K. Schulten (Eds.), *Models of Neural Networks III*. New York: Springer-Verlag.
- MacKay, D. J. C. (1995a). Bayesian non-linear modelling for the 1993 energy prediction competition. In G. Heidbreder (Ed.), *Maximum Entropy and Bayesian Methods, Santa Barbara 1993*. Dordrecht: Kluwer.
- MacKay, D. J. C. (1995b). Developments in probabilistic modelling with neural networks—ensemble learning. In B. Kappen and S. Gielen (Eds.), *Neural Networks: Artificial Intelligence and Industrial Applications. Proceedings of the 3rd Annual Symposium on Neural Networks, Nijmegen, Netherlands, 14-15 September 1995*, Berlin, pp. 191–198. Springer.

- MacKay, D. J. C. (1997). Gaussian processes - a replacement for supervised neural networks? Lecture notes for a tutorial at NIPS 1997. <ftp://wol.ra.phy.cam.ac.uk/pub/mackay/gp.ps.gz>.
- MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In L. M. Le Cam and J. Neyman (Eds.), *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, Volume 1, Berkeley, California, pp. 281–297. University of California Press.
- Madsen, H. (1989). *Tidsrækkeanalyse* (Draft ed.). Building 321, DTU, DK-2800 Lyngby, Denmark: Department of Mathematical Modelling, Technical University of Denmark. In danish.
- Madsen, P. L. (2000). Mysteriet om de forsvundne kulstofskelletter. In G. M. Knudsen, S. Vorstrup, G. Waldemar, and P. S. Sørensen (Eds.), *Festskrift i anledning af Olaf B. Paulsons 60 års fødselsdag*, pp. 17–20. In Danish.
- Madsen, P. L., N. F. Cruz, L. Sokoloff, and G. A. Dienel (1999, April). Cerebral oxygen/glucose ratio is low during sensory stimulation and rises above normal during recovery: excess glucose consumption during stimulation is not accounted for by lactate efflux from or accumulation in brain tissue. *Journal of Cerebral Blood Flow and Metabolism* **19**: 393–400. PMID: 10197509.
- Maes, F., A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens (1997, April). Multimodality image registration by maximization of mutual information. *IEEE Transactions of Medical Imaging* **16**(2): 187–198. PMID: 9101328.
- Maess, B., S. Koelsch, T. C. Gunter, and A. D. Friderici (2001, May). Musical syntax is processed in Broca's area: an MEG study. *Nature Neuroscience* **4**(5): 540–545. PMID: 11319564.
- Maguire, E. A., D. G. Gadian, I. S. Johnsrude, C. D. Good, J. Ashburner, R. S. J. Frackowiak, and C. D. Frith (2000, April). Navigation-related structural change in the hippocampi of taxi drivers. *Proceedings of the National Academy of Sciences USA* **97**(8): 4398–4403. PMID: 10716738. <http://www.pnas.org/cgi/pmidlookup?view=full&pmid=10716738>.
- Maintz, J. B. A. and M. A. Viergever (1998). A survey of medical image registration. *Medical Image Analysis* **2**(1): 1–36. <http://www.cs.uu.nl/people/twan/personal/media97.pdf>.
- Maisog, J. M. (2000, September). to (spatially) smooth or not to smooth raw fMRI data? SPM mailing list. <http://www.jiscmail.ac.uk/cgi-bin/wa.exe?A2=ind0009&L=spm&D=0&P=7404>.
- Makeig, S., T.-P. Jung, M. Westerfield, and T. J. Sejnowski (2001, June). Imaging event-related brain dynamics. *NeuroImage* **13**(6, part 2): S1309. <http://www.apnet.com/www/journal/hbm2001/11612.html>.
- Maki, A., Y. Yamashita, U. Ito, E. Watanabe, Y. Mayanagi, and H. Koizumi (2001, December). Spatial and temporal analysis of human motor activity using noninvasive nir topography. *Med. Phys.* **22**(12): 1997–2005.
- Malonek, D. and A. Grinvald (1996). Interactions between electrical activity and cortical microcirculation revealed by imaging spectroscopy; implications for functional brain imaging. *Science* **272**: 551–554.
- Marchini, J. L. and B. D. Ripley (2000, June). A new statistical approach to detecting significant activation in functional MRI. *NeuroImage* **11**(5): S609.
- Mardia, K. V., J. T. Kent, and J. M. Bibby (1979). *Multivariate Analysis*. London: Academic Press. ISBN 0124712525.
- Marquardt, D. W. (1963). An algorithm for least-squares estimation of non-linear parameters. *Journal of the Society of Industrial and Applied Mathematics* **11**(2): 431–441.
- Marrelec, G. and H. Benali (2001, June). Non-parametric Bayesian deconvolution of fMRI hemodynamic response function using smooth prior. *NeuroImage* **13**(6 (part 2)): S194. <http://www.apnet.com/www/journal/hbm2001/10906.html>.
- Marriott, F. H. C. (1952). Tests of significance in canonical analysis. *Biometrika* **39**: 58–64.
- Marriott, F. H. C. (1971, September). Practical problems in a method of cluster analysis. *Biometrics* **27**: 501–514.
- Martens, H. and T. Næs (1989). *Multivariate Calibration*. John Wiley & Sons. ISBN 0-471-93047-4.
- Martin, A., C. L. Wiggs, L. G. Ungerleider, and J. V. Haxby (1996, February). Neural correlates of category-specific knowledge. *Nature* **379**(6566): 649–652. PMID: 8628399.
- Martin, E., T. Thiel, F. Girard, and V. L. Marcar (2000, May). The effects of pentobarbital sedation on the cortical activation during visual stimulation in man (revisited). See Fox and Lancaster (2000), pp. S732. ISSN 1053-8119.
- Martínez, A. M., F. Hassainia, D. M. Mata, J. M. R. Medina, J. A. Leehan, and V. M. Bañuelos (2000). Remote brain mapping invocation. In *Proceedings of the 19th Annual International Conference of the IEEE Engineering in Medicine and Biology Society: Oct.30-Nov.2, 1997, Chicago II USA*, Volume 3, pp. 2339–2341. ISSN 0589-1019.

- Martinez, D. (1998, March). Neural tree density estimation for novelty detection. *IEEE Transactions on Neural Networks* **9**(2): 330–338. ISSN 1045-9227.
- Massy, W. F. (1965). Principal component analysis in exploratory data research. *Journal of the American Statistical Association* **60**: 234–256.
- Matthews, J., J. Ashburner, D. Bailey, R. Harte, P. Price, T. Jones, and P. A. Moonier (1995). The direct calculation of parametric images from raw PET data using maximum likelihood iterative reconstruction. In *1995 IEEE Nuclear Science Symposium and Medical Imaging Conference*, Volume 2, pp. 1311–1315. IEEE.
- MayerKress, G., C. Barcys, and W. Freeman (1991). Attractor reconstruction from event-related multi-electrode eeg data. In *Mathematical approaches to brain functioning diagnostics*. New York: Manchester University Press.
- Mazziotta, J. C. and R. S. J. Frackowiak (2000, May). The study of human disease with brain mapping methods. Chapter 1, pp. 3–31. San Diego, California: Academic Press. ISBN 0124814603.
- Mazziotta, J. C., A. W. Toga, A. C. Evans, P. T. Fox, and J. L. Lancaster (1995, June). A probabilistic atlas of the human brain: Theory and rationale for its development. the international consortium for brain mapping (icbm). *NeuroImage* **2**(2a): 89–101.
- McCain, K. W. (1990, September). Mapping authors in intellectual space: A technical overview. *Journal of the American Society for Information Science* **41**(6): 433–443.
- McCain, K. W. (1998, March-April). Neural networks research in context: A longitudinal journal cocitation analysis of an emerging interdisciplinary field. *Scientometrics* **41**(3): 389–410.
- McColl, J. H., A. P. Holmes, and I. Ford (1994). Statistical methods in neuroimaging with particular application to emission tomography. *Statistical Methods in Medical Research* **3**(1): 63–86.
- McCorry, S. J. and I. Ford (1991, November). Multivariate analysis of spect images with illustrations in Alzheimer's disease. *Stat. Med.* **10**(11): 1711–1718. PMID: 1792465.
- McDowall, I. E., M. T. Bolas, S. D. Pieper, S. S. Fisher, and J. Humphries (1990, September). Implementation and integration of a counterbalanced CRT-based stereoscopic display for interactive viewpoint control in virtual environment applications. In S. S. Fisher and J. O. Merritt (Eds.), *Proceedings of the 1990 SPIE Conference on Stereoscopic Displays and Applications, Santa Clara, California*, Volume 1256, pp. 136–146. The International Society for Optical Engineering. ISBN 0-8194-0303-2.
- McGonigle, D. J., A. M. Howseman, B. S. Athwal, K. J. Friston, R. S. J. Frackowiak, and A. P. Holmes (2000, June). Variability in fmri: An examination of intersession differences. *NeuroImage* **11**(6): 708–734.
- McGuire, P. K., E. Paulesu, R. S. J. Frackowiak, and C. D. Frith (1996, September). Brain activity during stimulus independent thought. *NeuroReport* **7**(13): 2095–2099. PMID: 8930966.
- McIntosh, A. R., F. L. Bookstein, J. V. Haxby, and C. L. Grady (1996, June). Spatial pattern analysis of functional brain images using Partial Least Square. *NeuroImage* **3**(3 part 1): 143–157. PMID: 9345485. ftp://ftp.rotman-baycrest.on.ca/pub/Randy/PLS/pls_article.pdf.
- McIntosh, A. R., L. Nyberg, F. L. Bookstein, and E. T. Tulving (1997). Differential functional connectivity of prefrontal and medial temporal cortices during episodic memory retrieval. *Human Brain Mapping* **5**(4): 323–327.
- McKeown, M. J., T.-P. Jung, S. Makeig, G. Brown, S. S. Kindermann, T.-W. Lee, and T. J. Sejnowski (1998a, February). Spatially independent activity patterns in functional MRI data during the Stroop color-naming task. *Proceedings of the National Academy of Sciences USA* **95**: 803–810. <file://ftp.cnl.salk.edu/pub/jung/PNASfMRI.ps.Z>.
- McKeown, M. J., S. Makeig, G. B. Brown, T.-B. Jung, S. S. Kindermann, A. J. Bell, and T. J. Sejnowski (1998b). Analysis of fMRI data by blind separation into independent spatial components. *Human Brain Mapping* **6**: 160–188.
- McLachlan, G. J. (1982). The classification and mixture maximum likelihood approaches to cluster analysis. Volume 2, Chapter 9, pp. 199–208. Amsterdam: North-Holland Publishing Company.
- Mencl, W. E., K. R. Pugh, S. E. Shaywitz, B. A. Shaywitz, R. K. Fullbright, R. T. Constable, P. Skudlarski, L. Katz, K. E. Marchione, C. Lacadie, and J. C. Gore (2000, October). Network analysis of brain activations in working memory: behavior and age relationships. *Microscopy Research and Technique* **51**(1): 64–74. ISSN 1097-0029. PMID: 11002354. <http://www3.interscience.wiley.com/cgi-bin/abstract/73500991/START>.

- Mendelzon, A. and D. Rafiei (2000, September). What do the neighbours think? computing web page reputations. *Bulletin of the Technical Committee on Data Engineering* **23**(3): 9–16. <http://www.research.microsoft.com/research/db/debull/A00sept/mendelz.ps>. ResearchIndex: <http://citeseer.nj.nec.com/374160.html>.
- Menon, R. S., D. C. Luknowsky, and J. S. Gati (1998, September). Mental chronometry using latency-resolved functional MRI. *Proceedings of the National Academy of Sciences USA* **95**(18): 10902–10907. PMID: 9724802. <http://www.pnas.org/cgi/content/full/95/18/10902>.
- Menon, R. S., S. Ogawa, X. Hu, J. P. Strupp, P. Anderson, and K. Ugurbil (1995, March). BOLD based functional MRI at 4 Tesla includes a capillary bed contribution: echo-planar imaging correlates with previous optical imaging using intrinsic signals. *Magnetic Resonance in Medicine* **33**(3): 453–459. PMID: 7760717.
- Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller (1953). Equations of state calculations by fast computing machine. *Journal of Chemical Physics* **21**: 1087–1091(1092?).
- Meyer, J. H., R. N. Gunn, R. Myers, and P. M. Grasby (1999, May). Assessment of spatial normalization of PET ligand images using ligand specific templates. *NeuroImage* **9**(5): 545–553. PMID: 10329294.
- Meyer-Lindenberg, A. (1998, July). RE: Talairach vs. MNI space. The SPM email discussion list, <http://www.mailbase.ac.uk/lists/spm/1998-06/0079.html>.
- Miki, A., J. Raz, T. G. van Erp, C. S. Liu, J. C. Haselgrove, and G. T. Liu (2000, May). Reproducibility of visual activation in functional imaging and effects of postprocessing. *AJNR American Journal of Neuroradiology* **21**(5): 910–915. PMID: 10815667.
- Miller, D. J. (1998, March). *Prescript: Programme Structure and functional Description*. New Zealand: The New Zealand Digital Library, University Waikato. Distributed with Prescript available at <http://www.nzdl.org/html/prescript.html>.
- Mishkin, M., L. G. Ungerleider, and K. A. Macko (1983). Object vision and spatial vision: Two cortical pathways. *Trends in Neurosciences* **6**: 414–417.
- Mitchell, T. A. (Ed.) (1997, March). *Machine Learning*. McGraw-Hill Series in Computer Science. McGraw-Hill. ISBN 0070428077.
- Mitra, S. K. (2001). *Digital Signal Processing: A Computer-base Approach* (Second ed.). McGraw-Hill Series in Electrical and Computer Engineering. Boston: McGraw-Hill. ISBN 0-07-118175-X.
- Moeller, J. R. and S. C. Strother (1991). A regional covariance approach to the analysis of functional patterns in positron emission tomographic data. *Journal of Cerebral Blood Flow and Metabolism* **11**: A121–A135.
- Moeller, J. R., S. C. Strother, J. J. Sidtis, and D. A. Rottenberg (1987). Scaled subprofile model: A statistical approach to the analysis of functional patterns in positron emission tomographic data. *Journal of Cerebral Blood Flow and Metabolism* **7**: 649–658.
- Molgedey, L. and H. G. Schuster (1994, June). Separation of a mixture of independent signals using time delayed correlations. *Physical Review Letters* **72**(23): 3634–3637. PMID: 10056251. <http://www.theo-physik.uni-kiel.de/thesis/molgedey94.ps.gz>. ResearchIndex: <http://citeseer.nj.nec.com/molgedey94separation.html>.
- Møller, M. F. (1993). A scaled conjugate gradient algorithm for fast supervised learning. *Neural Networks* **6**: 525–533.
- Moody, J. E. (1991). Note on generalization, regularization and architecture selection in nonlinear learning systems. In *First IEEE-SP Workshop on Neural Networks in Signal Processing*, Los Alamitos, California, pp. 1–10. IEEE Computer Society Press.
- Moody, J. E. (1992). The *effective* number of parameters: An analysis of generalization and regularization in nonlinear learning systems. In J. E. Moody, S. J. Hanson, and R. P. Lippmann (Eds.), *Advances in Neural Information Processing Systems: Proceedings of the 1991 Conference*, San Mateo, CA, pp. 847–854. Morgan Kaufmann Publishers. <http://www.cse.ogi.edu>. NIPS-4.
- Mørch, N. J. S. (1998). *A Multivariate Approach to Functional Neuro Modeling*. Ph.D. thesis, Technical University of Denmark, Lyngby, Denmark. ISSN 0909–3192.
- Mørch, N. J. S., L. K. Hansen, I. Law, S. C. Strother, C. Svarer, B. Lautrup, U. Kjems, N. Lange, and O. B. Paulson (1996a). Generalization and the bias-variance trade-off in models of functional brain activation. Originally submitted to IEEE Transactions on Medical Imaging.

- Mørch, N. J. S., L. K. Hansen, S. C. Strother, I. Law, C. Svarer, B. Lautrup, J. R. Anderson, N. Lange, and O. B. Paulson (1996b, June). Generalization performance of nonlinear vs. linear models for [^{15}O]water PET functional activation studies. See Belliveau, Fox, Kennedy, Rosen, and Ungeleider (1996), pp. S258. ISSN 1053–8119.
- Mørch, N. J. S., L. K. Hansen, S. C. Strother, C. Svarer, D. A. Rottenberg, B. Lautrup, R. Savoy, and O. B. Paulson (1997). Nonlinear versus linear models in functional neuroimaging: Learning curves and generalization crossover. In J. Duncan and G. Gindi (Eds.), *Information Processing in Medical Imaging, 15th International Conference, IPMI'97, Poultny, Vermont, USA, June 1997, Proceedings*, Volume 1230 of *Lecture Notes in Computer Science*, Berlin, pp. 259–270. Springer-Verlag. <http://nru.dk/cgi-bin/extract?morchn1997a>.
- Mørch, N. J. S., U. Kjems, L. K. Hansen, C. Svarer, I. Law, B. Lautrup, S. C. Strother, and K. Rehm (1995). Visualization of neural networks using saliency maps. In *Proceedings of 1995 IEEE International Conference on Neural Networks*, Volume 4, pp. 2085–2090. ISBN . <http://isp.imm.dtu.dk/publications/1995/nmorch.icnn95.ps.Z>.
- Moreno, J. L. (1934). *Who shall survive? A new Approach to the problem of human relations*. Nervous and Mental Disease Publishing Co.
- Morosan, P., J. Rademacher, A. Schleicher, K. Amunts, T. Schormann, and K. Zilles (2001, April). Human primary auditory cortex: Cytoarchitectonic subdivisions and mapping into a spatial reference system. *NeuroImage* **13**(4): 684–701. PMID: 11305897.
- Morris, T. A. and K. W. McCain (1998, September-October). The structure of medical informatics journal literature. *Journal of the American Medical Informatics Association* **5**(5): 448–466. PMID: 9760393. <http://www.jamia.org/cgi/content/full/5/5/448>.
- Mosher, J. C., P. S. Lewis, and R. M. Leahy (1992). Multiple dipole modeling and localization from spatiotemporal MEG data. *IEEE Trans. Biod. Eng* **39**: 541–557.
- Mueller, K., T. Welsh, W. Zhu, J. Meade, and N. Volkow (2000, November). BrainMiner: A visualization tool for ROI-based discovery of functional relationships in the human brain. In *New Paradigms in Information Visualization and Manipulation (NPIMV)*, Washington DC. <http://www.cs.sunysb.edu/~mueller/papers/npivm.pdf>.
- Munk, O. L. (1999). Automated image registration of PET brain scans using neural networks. See The International Society for Cerebral Blood Flow and Metabolism (1999), pp. S782. Brain '99 and BrainPET '99, XIXth Interation Symposium on Cerebral Blood Flow, Metabolism and Function, IVth Interation Conference on Quantification of Brain Functional with PET.
- Munzner, T. and P. Burchard (1995, December). Visualizing the structure of the world wide web in 3d hyperbolic space. In *Proceedings of VRML'95, December 14–15 1995*, pp. 33–38. ACM SIGGRAPH. <http://graphics.Stanford.EDU/papers/webviz/webviz.ps.gz>.
- Murata, N., S. Yoshizawa, and S. Amari (1991). A criterion for determining the number of parameters in an artificial neural network model. In *Artificial Neural Networks Proceedings of ICANN-91*, Volume 1, Amsterdam, pp. 9–14. North Holland.
- Murata, N., S. Yoshizawa, and S. Amari (1993). Learning curves, model selection and complexity of neural networks. See Hanson, Cowan, and Lee Giles (1993), pp. 607–614. NIPS-5.
- Murata, N., S. Yoshizawa, and S. Amari (1994). Network information criterion — determining the number of hidden units for an artificial neural network model. *IEEE Transaction on Neural Networks* **5**(6): 865–872.
- Murtagh, B. A. and R. W. H. Sargent (1969). A constrained minimization method with quadratic convergence. In R. Fletcher (Ed.), *Optimization*. London: Academic Press.
- Myers, R., R. N. Gunn, V. J. Cunningham, R. B. Banati, and T. Jones (1999). Cluster analysis and the reference tissue model in the analysis of clinical [^{11}C]PK11195 PET. See The International Society for Cerebral Blood Flow and Metabolism (1999), pp. S789. Brain '99 and BrainPET '99, XIXth Interation Symposium on Cerebral Blood Flow, Metabolism and Function, IVth Interation Conference on Quantification of Brain Functional with PET.
- Najork, M. A. and M. H. Brown (1995, June). Obliq-3D: A high-level, fast-turnaround 3D animation system. *IEEE Transactions on Visualization and Computer Graphics* **1**(2): 175–193. <http://www.research.compaq.com/SRC/3D-animate/tvcg/tvcg.html>.
- Nanba, H., N. Kando, and M. Okumura (2000). Classification of research papers using citation links and citation types: Towards automatic review article generation. In *The 11th SIG Classification Research Workshop, Classification for User Support and Learning, 2000.11, in Chicago, USA*, pp. 117–134. The American Society for Information Science (ASIS). http://galaga.jaist.ac.jp:8000/~nanba/study/asis2000_1.ps.gz.

- Neal, R. M. (1996). *Bayesian Learning for Neural Networks*. Number 118 in Lecture Notes in Statistics. New York, USA: Springer. ISBN 0-387-94724-8.
- Nelder, J. A. and R. Mead (1965). A simplex method for function minimization. *Computer Journal* 7: 308-313.
- Netsiri, C., S. Gustard, T. A. Carpenter, E. J. Williams, and C. L.-H. Huang (2000, May). Reliability of blind separation of non-smoothed fMRI data using ICA. See Fox and Lancaster (2000). ISSN 1053-8119. <http://www.academicpress.com/www/journal/hbm2000/6392.html>.
- Neufang, M., H. Obrig, R. R. Wenzel, M. Kohl, U. Scholz, J. Malak, and A. Villringer (1999, June). Spontaneous low frequency oscillations are attenuated by visual stimulation: A NIRS study. See Rosen, Seitz, and Volkmann (1999), pp. S311. ISSN 1053-8119.
- Nevill-Manning, C. C., T. Reed, and I. H. Witten (1998). Extracting text from PostScript. *Software — Practice and Experience* 28(5): 481-491. <http://www.cs.waikato.ac.nz/~ihw/papers/98NM-Reed-IHW-Extract-Text.pdf>.
- Nielsen, F. Å. (1996a, February). Visualization and analysis of 3D functional brain images. Master's thesis, Department of Mathematical Modelling, Technical University of Denmark, Lyngby, Denmark. <http://eivind.imm.dtu.dk>.
- Nielsen, F. Å. (1998, November). *polyr; Polygon generation program* (1.0.0 ed.). Lyngby, Denmark: Section for Digital Signal Processing, Department of Mathematical Modelling, Technical University of Denmark. <http://hendrix.imm.dtu.dk/software/polyr/doc/polyrdoc.ps.gz>.
- Nielsen, F. Å., C. Goutte, and L. K. Hansen (1999, June). Artificial neural network model for fMRI timeseries and a framework for comparison of convolution models. See Rosen, Seitz, and Volkmann (1999), pp. S173. ISSN 1053-8119. <http://eivind.imm.dtu.dk/publications/1999/nielsen.hbm99.ps.gz>.
- Nielsen, F. Å., C. Goutte, and L. K. Hansen (2001a, June). Modelling the fMRI response using smooth FIR filters. *NeuroImage* 13(6, part 2): S210. <http://www.apnet.com/www/journal/hbm2001/10284.html>, <http://hendrix.imm.dtu.dk/staff/fnielsen/ps/Nielsen2001Modelling.ps.gz>.
- Nielsen, F. Å. and L. K. Hansen (1997, November). Interactive information visualization in neuroimaging. See Ebert and Nicholas (1997). ISBN 1581130511. <http://hendrix.imm.dtu.dk/staff/fnielsen/ps/Nielsen1997Interactive.ps.gz>.
- Nielsen, F. Å. and L. K. Hansen (1998, May). Neuroinformatics based on VRML. See Paus, Gjedde, and Evans (1998), pp. S782. ISSN 1053-8119. <http://eivind.imm.dtu.dk/publications/1998/Nielsen1998Neuroinformatics.ps.gz>.
- Nielsen, F. Å. and L. K. Hansen (1999, May). Modeling of BrainMap data. <http://eivind.imm.dtu.dk/publications/1999/nielsen.nips99.ps.gz>. ResearchIndex: <http://citeseer.nj.nec.com/367960.html>. Submitted to but not accepted for Neural Information Processing Systems 12.
- Nielsen, F. Å. and L. K. Hansen (2000a, April). Experiences with Matlab and VRML in functional neuroimaging visualizations. In S. Klasky and S. Thorpe (Eds.), *Visualization Development Environments (VDE2000)*, April 27 - April 28, 2000, Princeton Plasma Physics Laboratory, Princeton, New Jersey. <http://w3.pppl.gov/vde2000/e proceedings.html>. ResearchIndex: <http://citeseer.nj.nec.com/309470.html>.
- Nielsen, F. Å. and L. K. Hansen (2000b, January). Functional volumes modeling using kernel density estimation. <http://hendrix.imm.dtu.dk/staff/fnielsen/ps/Nielsen2000Functional.ps.gz>. ResearchIndex: <http://citeseer.nj.nec.com/did/281332>. Submitted to *Sixth International Conference on Functional Mapping of the Human Brain* but lost in the submission database.
- Nielsen, F. Å. and L. K. Hansen (2001a, June). Author cocitation analysis of articles from "NeuroImage". *NeuroImage* 13(6, part 2): S212. <http://www.apnet.com/www/journal/hbm2001/10833.html>, <http://hendrix.imm.dtu.dk/staff/fnielsen/ps/Nielsen2001Author.ps.gz>. ResearchIndex: <http://citeseer.nj.nec.com/402733.html>.
- Nielsen, F. Å. and L. K. Hansen (2001b, March). Modeling of activation data in the BrainMap™ database: Detection of outliers. http://www.imm.dtu.dk/~ps/Nielsen2000Modeling_text.ps. Submitted to Human Brain Mapping.
- Nielsen, F. Å., L. K. Hansen, and U. Kjems (2000, May). Mining the BrainMap™ database: Detection of outliers. <http://hendrix.imm.dtu.dk/staff/fnielsen/ps/Nielsen2000Data.ps.gz>. ResearchIndex: <http://citeseer.nj.nec.com/nielsen00mining.html>. Submitted to NIPS*2000 and not accepted.
- Nielsen, F. Å., L. K. Hansen, and U. Kjems (2001b, June). Modeling of locations in the BrainMap database: Detection of outliers. *NeuroImage* 13(6, part 2): S211. http://www.imm.dtu.dk/~fn/ps/Nielsen2001Modeling_abstract.ps.gz. ResearchIndex: <http://citeseer.nj.nec.com/406116.html>.

- Nielsen, F. Å., L. K. Hansen, and S. C. Strother (1998, May). Canonical ridge analysis with ridge parameter optimization. See Paus, Gjedde, and Evans (1998), pp. S758. ISSN 1053–8119. <http://eivind.imm.dtu.dk/publications/1998/Nielsen1998Canonical.ps.gz>.
- Nielsen, F. Å., L. K. Hansen, P. Toft, C. Goutte, N. J. S. Mørch, C. Svarer, R. Savoy, B. R. Rosen, E. Rostrup, and P. Born (1997, May). Comparison of two convolution models for fMRI time series. See Friberg, Gjedde, Holm, Lassen, and Nowak (1997), pp. S473. ISSN 1053–8119. <http://eivind.imm.dtu.dk/publications/1997/HBM97.compare.poster473.ps.gz>.
- Nielsen, H. B. (1999). Damping parameter in Marquardt's method. Technical Report IMM-REP-1999-05, Department of Mathematical Modelling, Technical University of Denmark, Lyngby, Denmark. <http://www.imm.dtu.dk/documents/users/hbn/publ/TR9905.ps.Z>.
- Nielsen, J. (1996b, May). Top ten mistakes in web design. Internet. <http://www.useit.com/alertbox/9605.html>. Alertbox.
- Nielsen, J. (1997, December). Changes in web usability since 1994. <http://www.useit.com/alertbox/9712a.html>.
- Noll, D. C., C. R. Genovese, A. L. Vazquez, and W. Eddy (1996). Evaluation of respiratory artifact correction techniques in fMRI using ROC analysis. In *Proceedings of 4th Meeting of the International Society of Magnetic Resonance*, pp. 343.
- Nooruddin, F. S. and G. Turk (1999). Simplification and repair of polygonal models using volumetric techniques. Technical Report GITGVU -99-37, Georgia Institute of Technology. ResearchIndex: <http://citeseer.nj.nec.com/nooruddin99simplification.html>.
- North, S. C. (1992, October). *NEATO User's Guide*. Murray Hill, NJ: AT&T Bell Laboratories.
- Northoff, G., A. Richter, M. Gessner, F. Schlagenhaut, J. Fell, F. Baumgart, T. Kaulisch, R. Kotter, K. E. Stephan, A. Leschinger, T. Hagner, B. Bargel, T. Witzel, H. Hinrichs, B. Bogerts, H. Scheich, and H. J. Heinze (2000, January). Functional dissociation between medial and lateral prefrontal cortical spatiotemporal activation in negative and positive emotions: a combined fMRI/MEG study. *Cerebral Cortex* **10**(1): 93–107. PMID: 10639399.
- Northway, M. L. (1940). A method for depicting social relationships obtained by sociometric testing. *Sociometry* **3**: 144–150.
- Nowinski, W. L., R. N. Bryan, and R. Raghavan (1997). *The Electronic Clinical Brain Atlas*. Thieme. ISBN 3131076615. <http://www.thieme.de/detailseiten/3131076615.html>. CD-ROM.
- Ogawa, S., T. M. Lee, A. R. Kay, and D. W. Tank (1990, Dec). Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proceedings of the National Academy of Sciences of the United States of America* **87**(24): 9868–9872.
- Ogawa, S., D. W. Tank, R. Menon, J. Ellermann, S.-G. Kim, H. Merkle, and K. Ugurbil (1992, July). Intrinsic signal changes accompanying sensory stimulation — functional brain mapping with magnetic-resonance-imaging. *Proceedings of the National Academy of Science of the United States of America* **89**(13): 5951–5955.
- Ohlsson, M., C. Peterson, H. Pi, T. Rögnvaldsson, and B. Söderberg (1994). Predicting system loads with artificial neural networks — methods and results from “the Great Energy Predictor Shootout”. In *1994 Annual Proceedings of the American Society of Heating, Refrigerating and Air-Conditioning Engineers, Inc.*, Volume 100, pp. 1063–1074.
- Ollinger, J. M. and M. P. McAvoy (2000, May). A homogeneity correction for post-hoc ANOVAs in FMRI. *NeuroImage* **11**(5, part 2): S604. <http://www.apnet.com/www/journal/hbm2000/7128.html>.
- Ollinger, J. M., G. L. Shulman, and M. Corbetta (2001a, January). Separating processes within a trial in event-related functional MRI. I. analysis. *NeuroImage* **13**(1): 218–229. PMID: 11133324.
- Ollinger, J. M., G. L. Shulman, and M. Corbetta (2001b, January). Separating processes within a trial in event-related functional MRI. I. the method. *NeuroImage* **13**(1): 210–217. PMID: 11133323.
- Ono, M., S. Kubik, and C. D. Abernathy (1990). *Atlas of the Cerebral Sulci*. Stuttgart: Georg Thieme Verlag. ISBN 3–13–732101–8.
- Østergaard, L., R. M. Weiskoff, D. A. Chesler, C. Gyldensted, and B. R. Rosen (1996, November). High resolution measurement of cerebral blood flow using intravascular tracer bolus passages. part I: Mathematical approach and statistical analysis. *Magnetic Resonance in Medicine* **36**(5): 715–725. PMID: 8916022.
- O'Sullivan, B. T., P. E. Roland, and R. Kawashima (1994, January). A PET study of somatosensory discrimination in man. Microgeometry versus macrogeometry. *European Journal of Neuroscience* **6**(1): 137–148. PMID: 8130929.

- Owen, A. M., B. Milner, M. Petrides, and A. C. Evans (1996, August). Memory for object features versus memory for object location: A positron-emission tomography study of encoding and retrieval processes. *Proceedings of the National Academy of Sciences USA* **93**(17): 9212–9217. PMID: 8799180.
- Page, L. (1997). PageRank: Bringing order to the web. Stanford Digital Libraries Working Paper 1997-0072, Stanford University. <http://hci.stanford.edu/~page/papers/pagerank/>.
- Palubinskas, G., X. Descombes, and F. Kruggel (1998). An unsupervised clustering method using the entropy minimization. In A. K. Jain, S. Venkatesh, and B. C. Lovell (Eds.), *Proceedings. Fourteenth International Conference on Pattern Recognition*, Volume 2, pp. 1816–1818. IEEE Computer Society.
- Pardo, J. V. and P. T. Fox (1993). Preoperative assessment of the cerebral hemispheric dominance for language with CBF PET. *Human Brain Mapping* **1**: 57–68.
- Pardo, J. V., M. E. Raichle, and P. T. Fox (1991, January). Localization of a human system for sustained attention by positron emission tomography. *Nature* **349**(6304): 61–63. PMID: 1985266.
- Park, T., H. Pan, Y. Yan, W. Engelen, D. Ruppert, D. A. Silbersweig, and E. Stern (2000, May). Correction of physiological fluctuation in fMRI using linear prediction. See Fox and Lancaster (2000), pp. S472. ISSN 1053-8119. <http://www.academicpress.com/www/journal/hbm2000/6037.html>.
- Parzen, E. (1962). On the estimation of a probability density function and mode. *Annals of Mathematical Statistics* **33**: 1065–1076.
- Pascual-Marqui, R. D., D. Lehmann, T. Joenig, K. Kochi, M. C. Merlo, D. Hell, and M. Koukko (1999, June). Low resolution brain electromagnetic tomography (LORETA) functional imaging in acute, neuroleptic-naive, first-episode, productive schizophrenia. *Psychiatry Res.* **90**(3): 169–179.
- Pastoor, S. and M. Wöpping (1997). 3-d displays: A review of current technologies. *DISPLAYS* **17**: 100–110. <http://atwww.hhi.de:80/~blick/Papers/displays97/displays97.html>.
- Pauling, L. and C. D. Coryell (1936). The magnetic properties and structure of hemoglobin, oxyhemoglobin and carbomonoxyhemoglobin. *Proceedings of the National Academy of Sciences, USA* **22**: 210–216.
- Paus, T. (1996, June). Location and function of the human frontal eye field. *Neuropsychologia* **34**(6): 475–483. 8736560.
- Paus, T., A. Gjedde, and A. C. Evans (Eds.) (1998, May). *Fourth International Conference on Functional Mapping of the Human Brain. NeuroImage*, Volume 7. Academic Press. ISSN 1053–8119.
- Pearce, S. C. (1982). Analysis of covariance. See Kotz, Johnson, and Read (1982), pp. 61–69. ISBN 0–471–05546–8.
- Pearson, H. (2001, June). Genetic nomenclature. *Nature* **411**(6838): 631–632.
- Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. *The London, Edinburgh and Dublin Philosophical Magazine and Journal of Science* **2**: 559–572.
- Pedersen, M. W. (1997, August). *Optimization of Recurrent Neural Networks for Time Series Modeling*. Ph.d. thesis, Department of Mathematical Modelling, Technical University of Denmark, Lyngby, Denmark.
- Pelizzari, C. A., G. T. Chen, D. R. Spelbring, R. R. Weichselbaum, and C. T. Chen (1989, January-February). Accurate three-dimensional registration of CT, PET, and/or MR images of the brain. *Journal of Computer Assisted Tomography* **13**(1): 20–26. PMID: 2492038.
- Peltier, S. J. (2000). Analysis of fMRI signal and noise component TE dependence. *NeuroImage* **11**(5, part 2): S623.
- Pennock, D. M., S. Lawrence, C. L. Giles, and F. Å. Nielsen (2001a, August). Extracting collective probabilistic forecasts from web games. In *KDD-2001, The Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, August 26 - 29, 2001, San Francisco, California, USA*. SIGKDD, Association for Computing Machinery. <http://www.neci.nec.com/~lawrence/papers/am-kdd01/am-kdd01.ps.gz>.
- Pennock, D. M., S. Lawrence, C. L. Giles, and F. Å. Nielsen (2001b, February). The power of play: Efficiency and forecast accuracy in web market games. Technical Report 2000-168, NEC Research Institute, 4 Independence Way, Princeton, New Jersey. <http://artificialmarkets.com/am/pennock-neci-tr-2000-168.ps>.
- Pennock, D. M., S. Lawrence, C. L. Giles, and F. Å. Nielsen (2001c, February). The real power of artificial markets. *Science* **291**(5506): 987–988.
- Penrose, R. (1955). A generalized inverse for matrices. *Proc. Cambridge Philos. Soc.* **51**: 406–413.

- Perani, F. L. D., E. Guigon, V. Bettinardi, M. Carrozzo, F. Grassi, Y. Rossetti, and F. Fazio (1997, February). Visuomotor transformations for reaching to memorized targets: a PET study. *NeuroImage* **5**(2): 129–146. PMID: 9345543.
- Persson, O. (2000, September). A tribute to Eugene Garfield — discovering the intellectual base of his discipline. *Current Science* **5**(10): 590–591. <http://tejas.serc.iisc.ernet.in/~currsci/sep102000/590.pdf>.
- Persson, O. (2001, August). BIBEXCEL, a tool-box for scientometric analysis. Internet. Visited 2000-08-13.
- Persson, O., T. Luukkonen, and S. Hälikkä (2000). A bibliometric study of Finnish science. Working Papers 48/00, VTT, Group for Technology Studies. ISSN 1239-0259. <http://www.vtt.fi/ttr/pdf/wp48.pdf>.
- Persson, O., P. Stern, and K.-G. Holmberg (1992). BIBMAP: a toolbox for mapping the structure of scientific literature. In *Representations of Science and Technology*, pp. 189–199. DSWO Press.
- Petersen, K. S. (2000, January). Signalseparation med uafhængig komponent analysis (ICA). Master's thesis, Department of Mathematical Modelling, Technical University of Denmark, Lyngby, Denmark. IMM-EKS-2000-3, In Danish.
- Petersen, N. V., J. L. Jensen, J. Burchhardt, and H. Stødkilde-Jørgensen (1998, May). State space models for physiological noise in fMRI time series. See Paus, Gjedde, and Evans (1998), pp. S592. ISSN 1053–8119.
- Petersen, S. E., P. T. Fox, M. I. Posner, M. Mintun, and M. E. Raichle (1988, February). Positron emission tomographic studies of the cortical anatomy of single-word processing. *Nature* **331**(6157): 585–589. PMID: 3277066.
- Petersen, S. E., P. T. Fox, A. Z. Snyder, and M. E. Raichle (1990, August). Activation of extrastriate and frontal cortical areas by visual words and word-like stimuli. *Science* **249**: 1041–1044.
- Petersson, K. M., T. E. Nichols, J.-B. Poline, and A. P. Holmes (1999a). Statistical limitations in functional neuroimaging. i. non-inferential methods and statistical models. *Philosophical Transactions of the Royal Society - Series B - Biological Sciences* **354**(1387): 1239–1260.
- Petersson, K. M., T. E. Nichols, J.-B. Poline, and A. P. Holmes (1999b). Statistical limitations in functional neuroimaging. ii. signal detection and statistical inference. *Philosophical Transactions of the Royal Society - Series B - Biological Sciences* **354**(1387): 1261–1282.
- Phelps, M. E. (1975). Application of annihilation coincidence detection to transaxial reconstruction tomography. *Nuclear Medicine* **16**: 210–224.
- Philips, D. L. (1962). A technique for the numerical solution of certain integral equations of the first kind. *J. ACM*: 84–97.
- Philipsen, P. A. and L. K. Hansen (1999). PET reconstruction with a Markov random field prior. See The International Society for Cerebral Blood Flow and Metabolism (1999), pp. S783. Brain '99 and BrainPET '99, XIXth Interation Symposium on Cerebral Blood Flow, Metabolism and Function, IVth Interation Conference on Quantification of Brain Functional with PET.
- Phong, B. T. (1975). Illumination for computer generated pictures. *Communications of the ACM* **18**(6): 311–317.
- Pi, H. and C. Peterson (1994, May). Finding the embedding dimension and variable dependencies in time series. *Neural Computation* **6**(3): 509–520.
- Picard, N. and P. L. Strick (1996, May/June). Motor areas of the medial wall: a review of their location and functional activation. *Cerebral Cortex* **6**(3): 342–353. PMID: 8670662.
- Pitkow, J. E. (1997, June). *Characterizing World Wide Web Ecologies*. Ph. D. thesis, Georgia Institute of Technology, Georgia, USA. <ftp://ftp.cc.gatech.edu/pub/gvu/tr/1997/97-16.ps.Z>.
- Pitkow, J. E. and P. L. Pirolli (2001, January). Method and apparatus for finding related documents in a collection of linked documents using a bibliographic coupling link analysis. United States Patent No. 6,182,091. http://www.delphion.com/details?pn=US06182091__, <http://164.195.100.11/netacgi/nph-Parser?d=PALL&p=1&u=/netahtml/srchnum.htm&r=1&f=G&l=50&s1='6,182,091'.WKU.&RS=PN/6,182,091>.
- Pizzi, N. J., R. A. Vivanco, and R. L. Somorjai (2001). Evident(tm): a functional magnetic resonance image analysis system. *Artificial Intelligence in Medicine* **21**(1–3): 263–269. <http://www.ee.umanitoba.ca/~pizzi/Papers/p008.pdf>.
- Polak, E. and G. Ribiere (1969). Note on the convergence of methods of conjugate directions. *Revue Francaise d'Informatique et de Recherche Operationnelle* **3**(16): 35–43.

- Poldrack, R. (2001, March). Re: Talairach. SPM mailing list. <http://www.jiscmail.ac.uk/cgi-bin/wa.exe?A2=ind0103&L=spm&D=0&P=28307>.
- Poline, J.-B. (1999a, March). `spm_clusters.m/c`. Distributed with SPM99.
- Poline, J.-B. (1999b, May). `spm_max.c`. Distributed with SPM99.
- Poline, J.-B., A. P. Holmes, K. J. Worsley, and K. J. Friston (1997, May). Statistical inference and the theory of random fields. In *SPMcourse, Short course notes*, Chapter 4. Institute of Neurology, Wellcome Department of Cognitive Neurology.
- Poline, J.-B. and B. M. Mazoyer (1994, July). Enhanced detection in brain activation maps using a multifiltering approach. *Journal of Cerebral Blood Flow and Metabolism* **14**(4): 639–642. PMID: 8014211.
- Poline, J.-B., R. Vandenberghe, A. P. Holmes, K. J. Friston, and R. S. J. Frackowiak (1996, August). Reproducibility of PET activation studies: Lessons from a multi-center european experiment. *NeuroImage* **4**(1): 24–54. PMID: 9345495.
- Poline, J.-B., K. J. Worsley, A. P. Holmes, R. S. J. Frackowiak, and K. J. Friston (1995, September-October). Estimating smoothness in statistical parametric maps: variability of p values. *Journal of Computed Assisted Tomography* **19**(5): 788–796.
- Poljak, B. T. (1964). Some methods of speeding up the convergence of iteration methods. *Z. VyCisl. Mat. i Mat. Fiz* **4**: 1–17.
- Pomerleau, D. A. (1993). Input reconstruction reliability estimation. See Hanson, Cowan, and Lee Giles (1993), pp. 279–286. NIPS-5.
- Porter, M. F. (1980). An algorithm for suffix stripping. *Program* **14**(3): 130–137.
- Posner, M. I. and M. E. Raichle (1995). *Précis of images of mind*. *Behavioral and Brain Sciences* **18**: 327–383.
- Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery (1992). *Numerical Recipes in C: The Art of Scientific Computing* (Second ed.). Cambridge, England: Cambridge University Press. ISBN 0521341085.
- Price, C. J. and K. J. Friston (1997, May). Cognitive conjunction: A new approach to brain activation experiments. *NeuroImage* **4**(5): 261–270. PMID: 9345555. <http://www.idealibrary.com/links/doi/10.1006/nimg.1997.0269>.
- Price, C. J. and K. J. Friston (1999). Scanning patients with tasks they can perform. *Human Brain Mapping* **8**(2–3): 102–108. PMID: 10524600.
- Price, C. J., D. J. Veltman, J. Ashburner, O. Josephs, and K. J. Friston (1999). The critical relationship between the timing of stimulus presentation and data acquisition in blocked designs with fMRI. *NeuroImage* **10**: 36–44.
- Prichard, J., D. Rothman, E. Novotny, O. Petroff, T. Kuwabara, M. Avison, A. Howseman, C. Hanstock, and R. Shulman (1991, July). Lactate rise detected by ¹H NMR in human visual cortex during physiologic stimulation. *Proceedings of the National Academy of Sciences of USA* **88**(13): 5829–5831. PMID: 2062861.
- Proakis, J. G. and D. G. Manolakis (1996). *Digital Signal Processing. Principles, Algorithms, and Applications* (Third ed.). Prentice Hall International.
- Purcell, E. M., H. C. Torrey, and R. V. Pound (1946). Resonance absorption by nuclear magnetic moments in solids. *Physical Review* **69**: 37–83.
- Purdon, P. L. and R. M. Weisskoff (1998). Effect of temporal autocorrelation due to physiological noise and stimulus paradigm on voxel-level false-positive rates in fMRI. *Human Brain Mapping* **6**(4): 239–249. PMID: 9704263.
- Purushotham, A., F. Å. Nielsen, L. K. Hansen, and S.-G. Kim (1999, June). Separation of motor preparation and execution regions using meta-K-means clustering on fMRI single trial data. See Rosen, Seitz, and Volkmann (1999), pp. S51. ISSN 1053–8119.
- Purushotham, A., F. Å. Nielsen, L. K. Hansen, and S.-G. Kim (2000, April). Meta-K-Means clustering as a means to separate different functional components of a cognitive task in fMRI single trial data. In *8th Scientific Meeting & Exhibition*. International Society for Magnetic Resonance in Medicine.
- Quinton, O., S. Gicquel, C. Tzourio, N. Mazoyer, M. Joliot, B. Mazoyer, and A. Alperovitch (1999). EVA833: a statistical MRI brain atlas based on 833 elderly subjects. In *Fifth International Conference on Functional Mapping of the Human Brain*, pp. S101.

- Rabi, I. I. and V. W. Cohen (1933). The nuclear spin of sodium. *Physical Review* **43**: 582.
- Rafiei, D. and A. O. Mendelzon (2000). What is this page known for? computing web page reputations. *Computer Networks* **33**: 823–835.
- Rajapakse, J. C., F. Kruggel, J. M. Maisog, and D. Y. von Cramon (1998). Modeling hemodynamic response for analysis of functional MRI time-series. *Human Brain Mapping* **6**(4): 283–300.
- Rao, A. R. and A. Jaimes (1999). Digital stereoscopic imaging. In J. O. Merritt (Ed.), *Stereoscopic Displays and Virtual Reality Systems VI*, Volume 3639 of *SPIE Proceedings*, Bellingham, Washington, pp. 144–154. The International Society for Optical Engineering. ISBN 0-8194-3110-9.
- Rao, C. R. (1964). The use and interpretation of principal component analysis in applied research. *Sankhyā A* **26**: 329–358.
- Rasmussen, C. E. (2000). The infinite gaussian mixture model. In *Advances in Neural Information Processing Systems 12*.
- Rasmussen, C. E. and Z. Ghahramani (2001). Occam's razor. See Leen, Dietterich, and Tresp (2001). <http://nips.djvuzone.org/djvu/nips13/RasmussenGhahramani.djvu>. NIPS-13.
- Rauch, S. L., M. A. Jenike, N. M. Alpert, H. C. R. Breiter, C. R. Savage, and A. J. Fischman (1994, January). Regional cerebral blood flow measured during symptom provocation in obsessive-compulsive disorder using oxygen 15-labeled carbon dioxide and positron emission tomography. *Archives of General Psychiatry* **51**(1): 62–70. PMID: 8279930.
- Ray, P. G., K. J. Meador, J. R. Smith, J. W. Wheless, M. Sittenfeld, and G. L. Clifton (1999, March). Physiology of perception: cortical stimulation and recording in humans. *Neurology* **52**(5): 1044–1049.
- Reber, A. S. (1995). *The Penguin Dictionary of Psychology*. Penguin.
- Rehm, K., K. Lakshminarayan, S. A. Frutiger, K. A. Schaper, D. L. Sumners, S. C. Strother, J. R. Anderson, and D. A. Rottenberg (1998, September). A symbolic environment for visualizing activated foci in functional neuroimaging datasets. *Medical Image Analysis* **2**(3): 215–226. PMID: 9873900. <http://www.sciencedirect.com/science/article/B6W6Y-45PJY0D-7/1/48196224354fdd62ea8c5a0d85379b07>.
- Rehm, K., K. A. Schaper, K. Lakshminarayan, and D. A. Rottenberg (2000). Symbolic representation of neuroimaging data: the corner cube environment. In S. Klasky and S. Thorpe (Eds.), *Visualization Development Environments (VDE2000)*, April 27 - April 28, 2000, Princeton Plasma Physics Laboratory, Princeton, New Jersey. <http://w3.pppl.gov/vde2000/eeproceedings/P19rehm.PDF>.
- Rehm, K., S. C. Strother, J. R. Anderson, K. A. Schaper, and D. A. Rottenberg (1994). Display of merged multimodality brain images using interleaved pixels with independent color scales. *Journal of Nuclear Medicine* **35**(11): 1815–1821. PMID: 7965164. <http://pet.med.va.gov:8080/papers/rehm-94a.html>.
- Richardson, S. and P. J. Green (1997). On bayesian analysis of mixtures with an unknown number of components (with discussion). *Journal of the Royal Statistical Society, Series B* **59**(4): 731–792.
- Richter, W., P. M. Andersen, A. P. Georgopoulos, and S.-G. Kim (1997, March). Sequential activity in human motor areas during a delayed cued finger movement task studied by time-resolved fMRI. *NeuroReport* **8**(5): 1257–1261.
- Ricotti, L. P., S. Ragazzani, and G. Martinelli (1988). Learning of word stress in a sub-optimal secondorder backpropagation neural network. In *Proceedings of the IEEE International Conferenec on Neural Networks*, Volume 1, San Diego, CA, pp. 355–361.
- Rinberg, D., H. Davidowitz, and N. Tishby (1999). Multi-electrode spike sorting by clustering transfer functions. See Kearns, Solla, and Cohen (1999), pp. 146–152. ISBN 0-262-11245-0. NIPS-11.
- Ripley, B. D. (1993). *Statistical aspect of neural networks*, Chapter 2, pp. 40–123. Number 50 in Monographs on Statistics and Applied Probability. London: Chapman & Hall.
- Ripley, B. D. (1996). *Pattern Recognition and Neural Networks*. Cambridge, United Kingdom: Cambridge University Press. ISBN 0521460867.
- Rissanen, J. (1978). Modeling by shortest data description. *Automatica* **14**: 465–471.
- Robb, R. A. (1999). 3-d visualization in biomedical applications. *Annu. Rev. Biomed. Eng.* **1**: 377–399.

- Roberts, S. and L. Tarassenko (1994). A probabilistic resource allocating network for novelty detection. *Neural Computation* **6**: 270–284.
- Roberts, S. J. (1999). Novelty detection using extreme value statistics. *IEE Proceedings-Vision, Image and Signal Processing* **146**(3): 124–129. ISSN 1350-245x.
- Robertson, G. G., J. D. Mackinlay, and S. K. Card (1991). Cone trees: Animated 3D visualizations of hierarchical information. In *Proceedings of CHI'91, ACM Conference on Human Factors in Computing Systems, New Orleans, Louisiana*, pp. 189–194. ACM.
- Robertson, G. G., J. D. Mackinlay, and S. K. Card (1994, March). Display of hierarchical three-dimensional structures with rotating substructures. US Patent: 5295243. <http://www.delphion.com/details?pn=US05295243>.
- Robertson, S. E. and K. Sparck Jones (1976). Relevance weighting of search terms. *Journal of the American Society of Information Science* **27**: 129–146.
- Robertson, S. E. and S. Walker (1994). Some simple effective approximation to the 2-Poisson model for probabilistic weighted retrieval. In W. B. Croft and C. J. van Rijsbergen (Eds.), *SIGIR 94, Proceedings of the Seventh International Conference on Research and Development in Information Retrieval*, pp. 232–241. Springer-Verlag.
- Robson, M. D., J. L. Dorosz, and J. C. Gore (1998). Measurements of the temporal fMRI response of the human cortex to trains of tones. *NeuroImage* **7**(3): 185–198. PMID: 9597660. Published erratum appears in *Neuroimage* 1998 Aug, **8**(2):228.
- Roland, P. E., J. Fredriksson, P. Svensson, K. Amunts, C. Cavada, R. Hari, A. Cowey, F. Crivello, S. Geyer, G. Kostopoulos, B. Mazoyer, D. Poppelwell, A. Schleicher, T. Schormann, M. Seppa, H. Uylings, K. de Vos, and K. Zilles (1999, June). ECHBD, a database for functional-structural and functional-functional relations in neuroimaging. In *Fifth International Conference on Functional Mapping of the Human Brain*, pp. S128.
- Roland, P. E., S. Geyer, K. Amunts, T. Schormann, A. Schleicher, A. Malikovic, and K. Zilles (1997, December). Cytoarchitectural maps of the human brain in standard anatomical space. *Human Brain Mapping* **5**(4): 222–227.
- Roland, P. E., C. J. Graufelds, J. Wåhlin, L. Ingelman, M. Andersson, A. Ledberg, J. Pedersen, S. Åkerman, A. Dabringhaus, and K. Zilles (1994). Human brain atlas: For high-resolution functional and anatomical mapping. *Human Brain Mapping* **1**: 173–184.
- Rorden, C. and M. Brett (2000, October). Reporting activity or lesion location. http://www.mrc-cbu.cam.ac.uk/Imaging/brodmann_areas.html.
- Rorden, C. and M. Brett (2001). Stereotaxic display of brain lesions. *Behavioural Neurology*. In press.
- Rosen, B. R., R. J. Seitz, and J. Volkmann (Eds.) (1999, June). *Fifth International Conference on Functional Mapping of the Human Brain, NeuroImage*, Volume 9. Academic Press. ISSN 1053–8119.
- Rosen, S. D., E. Paulesu, C. D. Frith, R. S. J. Frackowiak, G. J. Davies, T. Jones, and P. G. Camici (1994, July). Central nervous pathways mediating angina pectoris. *Lancet* **344**(8916): 147–150. PMID: 7912763.
- Rosenblatt, F. (1962). *Principals of Neurodynamics*. New York: Spartan.
- Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *The Annals of Mathematical Statistics* **1956**: 832–837.
- Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psych Bull.* **86**: 638–641.
- Rostrup, E., I. Law, M. Blinkenberg, H. B. Larsson, A. P. Born, S. Holm, and O. B. Paulson (2000, February). Regional differences in the CBF and BOLD responses to hypercapnia: a combined PET and fMRI study. *NeuroImage* **11**(2): 87–97. PMID: 10679182. <http://www.idealibrary.com/links/citation/1053-8119/11/87>.
- Rowe, J. B., I. Toni, O. Josephs, R. S. J. Frackowiak, and R. E. Passingham (2000, June). The prefrontal cortex: response selection or maintenance within working memory? *Science* **288**(5471): 1656–1660. PMID: 10834847.
- Roweis, S. (1998). EM algorithms for PCA and SPCA. In M. I. Jordan, M. J. Kearns, and S. A. Solla (Eds.), *Advances in Neural Information Processing Systems 10: Proceedings of the 1997 Conference*. MIT Press. ISBN 0262100762. <http://www.gatsby.ucl.ac.uk/~roweis/papers/empca.ps.gz>.
- Roweis, S. and C. Brody (1999). Linear heteroencoders. Technical Report GCNU-TR-1999-02, Gatsby Computational Neuroscience Unit, Alexandra House, 17 Queen Square, LONDON, WC1N 3AR, U.K.. <http://www.gatsby.ucl.ac.uk/~roweis/papers/lhet.ps.gz>.

- Roweis, S. and Z. Ghahramani (1999, February). A unifying review of linear gaussian models. *Neural Computation* **11**(2): 305–345.
- Roy, C. S. and C. S. Sherrington (1890). On the regulation of the blood supply of the brain. *J. Physiol. (London)* **11**: 85–108.
- Royackkers, N., M. Desvignes, H. Fawal, and M. Revenu (1999, December). Detection and statistical analysis of human cortical sulci. *NeuroImage* **10**(6): 625–641. <http://www.idealibrary.com/links/doi/10.1006/nimg.1999.0512>.
- Rubin, D. B. and D. T. Thayer (1982). EM algorithms for the ML factor analysis. *Psychometrika* **47**(1): 69–76.
- Saad, Z. S., E. A. Deyoe, and K. M. Ropella (1999, June). In the variance of activation delays: Methodological or physiological. *NeuroImage* **9**(6, part 2): S89.
- Saad, Z. S., K. M. Ropella, G. J. Carman, and E. A. DeYoe (1996, April/May). Temporal phase variation of fMR signals in vasculature versus parenchyma. In *Proceedings of the International Society for Magnetic Resonance in Medicine*, Volume 3, Berkely, California, pp. 1834. Society for Magnetic Resonance. ISSN 1065-9889.
- Sahami, M. (1999, December). *Using Machine Learning to Improve Information Access*. Ph.d. thesis, Computer Science Department, Stanford University, California. <http://www.sims.berkeley.edu/courses/is296a-3/f98/readings/sonia.ps>.
- Salton, G. (1989). *Automatic Text Processing: The Transformation, Analysis and Retrieval of Information by Computer*. Reading, Massachusetts: Addison-Wesley Publishing Company, Inc.
- Salton, G. and C. Buckley (1988). Term-weighting approaches in automatic text retrieval. *Information Processing and Management* **24**(5): 513–523.
- Salton, G. and C. Buckley (1989, April). On the automatic generation of content links in hypertext. Technical Report TR89–993, Department of Computer Science, Cornell University.
- Salton, G., A. Wong, and C. S. Yang (1975). A vector space model for automatic indexing. *Communication of the ACM* **18**: 613–620.
- Salton, G. and C. S. Yang (1973, December). On the specification of terms values in automatic indexing. *Journal of Documentation* **29**(4): 351–372.
- Sandor, S. and R. Leahy (1997, February). Surface-based labeling of cortical anatomy using deformable database. *IEEE Transactions on Medical Imaging* **16**(1): 41–54.
- Sarukkai, R. R. (2000). Link prediction and path analysis using Markov chains. *Computer networks* **33**: 377–386.
- Scarth, G. B., M. Alexander, M. McIntyre, B. Wowk, and R. L. Somorjai (1996). Artifact detection in fMRI using fuzzy clustering. In *Proceedings of 4th Meeting of the International Society of Magnetic Resonance*, pp. 1783.
- Scarth, G. B., M. McIntyre, B. Wowk, and R. L. Somorjai (1995, August). Detection of novelty in functional images using fuzzy clustering. In *Proceedings of the Society of Magnetic Resonance and the European Society For Magnetic Resonance In Medicine And Biology, Nice, France, August 19–25, 1995*, Volume 1, Berkely, California, pp. 238. Society of Magnetic Resonance. ISSN 1065-9889.
- Scarth, G. B. and R. L. Somorjai (2000, May). Method and apparatus for detection of events or novelties over a change of state. United States Patent No. 6,064,770.
- Schaltenbrand, G. and W. Wahren (1977, January). *Atlas for Stereotaxy of the Human Brain* (2nd ed.). Stuttgart: Thieme. ISBN 0865770557.
- Schaper, K. A., J. B. Arnold, J.-S. Liow, J. J. Stern, J. G. Sled, D. W. Shattuck, A. J. Worth, M. S. Cohen, R. M. Leahy, J. C. Mazziotta, and D. A. Rottenberg (2001, June). Evaluation of six algorithms for correcting intensity non-uniformity effects in mri volumes. *NeuroImage* **13**(6, part 2): S237. http://pet.med.va.gov:8080/papers/abstracts_posters/HBM2001/kschaper_HBM2001.pdf.
- Scharf, L. L. (1991). The SVD and reduced-rank signal processing. In R. J. Vaccaro (Ed.), *SVD and Signal Processing II: Algorithms Analysis and Applications*, pp. 3–31. Amsterdam, Netherlands: Elsevier.
- Schiessl, I., M. Stetter, J. E. W. Mayhew, S. Askew, N. McLoughlin, J. B. Levitt, J. S. Lund, and K. Obermayer (1999). Blind separation of spatial signal patterns from optical imaging records. In *ICA99 — International workshop on Independent Component Analysis and Blind Source Separation*, pp. 179–184. ftp://ftp.cs.tu-berlin.de/pub/local/ni/papers/schiessl_ICA_99.ps.gz.

- Schiessl, I., M. Stetter, J. E. W. Mayhew, N. McLoughlin, J. S. Lund, and K. Obermayer (2000, May). Blind signal separation from optical image recordings with extended spatial decorrelation. *IEEE Transaction on Biomedical Engineering* **47**(5): 573–577. ResearchIndex: <http://citeseer.nj.nec.com/266526.html>.
- Schlösser, R., M. Hutchinson, S. Joseffer, H. Rusinek, A. Saarimaki, J. Stevenson, S. L. Dewey, and J. D. Brodie (1998, April). Functional magnetic resonance imaging of human brain activity in a verbal fluency task. *Journal of Neurology, Neurosurgery, and Psychiatry* **64**(4): 492–495. PMID: 9576541. <http://jnnp.bmjjournals.com/cgi/content/full/64/4/492>.
- Schormann, T., S. Posse, S. Henn, and K. Zilles (1999, June). The new reference brain of the ECHB database. In *Fifth International Conference on Functional Mapping of the Human Brain, NeuroImage*, pp. S40.
- Schvaneveldt, R. W., F. T. Durso, and D. W. Dearholt (1989). Network structures in proximity data. In G. Bower (Ed.), *The Psychology of Learning and Motivation*, Volume 24, pp. 249–284. Academic Press.
- Schwartz, C. (1998). Web search engines. *Journal of the American Society for Information Science* **49**(11): 973–982.
- Schwarz, G. (1978). Estimating the dimension of a model. *Ann. Stat.* **14**: 461–64.
- Scroeder, W., K. Martin, and B. Lorensen (1997, December). *The Visualization Toolkit. An object-Oriented Approach to 3D Graphics* (Second ed.). Prentice Hall Computer Books. ISBN 0139546944.
- Seitz, R. J., C. Bohm, T. Greitz, P. E. Roland, L. Eriksson, H. Blomqvist, G. Rosenqvist, and B. Nordell (1990, July). Accuracy and precision of the computerized brain atlas programme for localization and quantification in positron emission tomography. *Journal of Cerebral Blood Flow and Metabolism* **10**(4): 443–457.
- Selden, C. R. (1992). Current bibliographies in medicine. Technical Report 92-13, National Library of Medicine, 8600 Rockville Pike, Bethesda, MD 20894. <http://www.nlm.nih.gov/pubs/cbm/metaanal.html>.
- Sen Gupta, A. (1982). Generalized canonical variables. See Kotz, Johnson, and Read (1982), pp. 326–330. ISBN 0–471–05546–8.
- Sergent, J., E. Zuck, M. Levesque, and B. MacDonald (1992, January/February). Positron emission tomography study of letter and object processing: empirical findings and methodological considerations. *Cerebral Cortex* **2**(1): 68–80. PMID: 1633409.
- Shanno, D. F. (1970). Conditioning of quasi-Newton method for function minimization. *Maths. Comput.* **24**: 27–30.
- Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal* **27**: 379–423, 623–656.
- Shao, J. (1993). Linear model selection by cross-validation. *Journal of the American Statistical Association* **88**(422): 486–494.
- Shaywitz, B. A., K. R. Pugh, R. T. Constable, S. E. Shaywitz, R. A. Bronen, R. K. Fulbright, D. P. Shankweiler, L. Katz, J. M. Fletcher, P. Skudlarski, and J. C. Gore (1995). Localization of semantic processing using functional magnetic resonance imaging. *Human Brain Mapping* **2**: 149–158.
- Shekhar, R., E. Fayyad, R. Yagel, and J. F. Cornhill (1996). Octree-based decimation of marching cubes surfaces. In *Proceedings of Visualization'96*, pp. 335–342. ResearchIndex: <http://www.researchindex.com/shekhar96octreebased.html>.
- Shepherd, G. M. (1994). *Neurobiology* (Third ed.). New York: Oxford University Press. ISBN 0195088433.
- Shepherd, G. M., J. S. Mirsky, M. D. Healy, M. S. Singer, E. Skoufos, M. S. Hines, P. M. Nadkarni, and P. L. Miller (1998, November). The Human Brain Project: neuroinformatics tools for integrating, searching and modeling multidisciplinary neuroscience data. *Trends in Neurosciences* **21**(11): 460–468.
- Shibata, R. (1976). Selection of the order of an autoregressive model by Akaike's information criterion. *Biometrika* **63**: 117–126.
- Shneiderman, B. (1991, March). Tree visualization with treemaps: a 2-d space-filling approach. Technical report, Human-Computer Interaction Lab, University of Maryland. <ftp://ftp.cs.umd.edu/pub/papers/papers/ncstrl.umcp/CS-TR-2645/CS-TR-2645.ps.Z>.
- Shneiderman, B. (1992, January). Tree visualization with treemaps: a 2-d space-filling approach. *ACM Transactions on Graphics* **11**(1): 92–99.
- Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman and Hall.

- Simon, T. J. (1999, October). The foundations of numerical thinking in a brain without numbers. *Trends in Cognitive Sciences* **3**(10): 363–364.
- Sjølund, A. (1965). *Gruppepsykologi*. Gyldendals Pædagogiske Bibliotek. Copenhagen, Denmark: Gyldendal. In Danish.
- Skerjanc, R. and S. Pastoor (1997, February). New generation of 3-d desktop computer interfaces. In *Proc. IS&T/SPIE EI'97 Conference*. http://atwww.hhi.de:80/~blick/Papers/New_generation_of_3-D_desktop/new_generation_of_3-d_desktop_.html.
- Small, H. G. (1973, July-August). Co-citation in the scientific literature: A new measure of the relationship between two documents. *Journal of the American Society for Information Science* **24**(4): 265–269.
- Small, H. G. and E. Garfield (1985). The geography of science: Disciplinary and national mappings. *Journal of Information Science* **11**: 147–159. <http://www.garfield.library.upenn.edu/essays/v9p325y1986.pdf>.
- Small, H. G. and M. E. D. Koenig (1977). Journal clustering using a bibliographic coupling method. *Information Processing and Management* **13**(5): 277–288.
- Smith, A. M., B. K. Lewis, U. R. Ruttiman, F. Q. Ye, T. M. Sinnweell, Y. Yang, J. H. Duyn, and J. A. Frank (1999, May). Investigation of low frequency drift in fmri signal. *NeuroImage* **9**(5): 526–533.
- Smith, J. R. and S.-F. Chang (1996, August). Searching for images and videos on the world-wide web. Technical Report 459-96-25, Department of Electrical Engineering and Center for Image Technology for New Media, Columbia University, New York. <ftp://ftp.ctr.columbia.edu/CTR-Research/advent/public/papers/96/smith96e.ps.gz>.
- Smith, S. (2000, May). Robust automated brain extraction. See Fox and Lancaster (2000), pp. S625. ISSN 1053-8119. <http://www.fmrib.ox.ac.uk/fsl/bet>.
- Smith, S., P. R. Bannister, C. Beckmann, M. Brady, S. Clare, D. Flitney, P. Hansen, M. Jenkinson, D. Lebovici, B. D. Ripley, M. Woolrich, and Y. Zhang (2001, June). FSL: New tools for functional and structural brain image analysis. *NeuroImage* **13**(6, part 2): S249. <http://www.academicpress.com/www/journal/hbm2001/10107.html>.
- Smyth, P. (1996, August). Clustering using monte carlo cross-validation. In *Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining*, pp. 126–133. AAAI Press. <http://www.ece.nwu.edu/~harsha/Clustering/kdd96.ps>. ResearchIndex: <http://citeseer.nj.nec.com/smyth96clustering.html>.
- Smyth, P. (2000). Model selection for probabilistic clustering using cross-validated likelihood. *Statistics and Computing* **10**(1): 63–72. http://www.ics.uci.edu/~datalab/papers/tr9809_rev.ps.gz.
- Snyder, H. and H. Rosenbaum (1999, September). Can search engines be used as tools for web-link analysis? a critical view. *Journal of Documentation* **55**(4): 375–384.
- Snyder, T. L. and J. M. Steele (1995). Probabilistic networks and network algorithms. In M. O. Ball, T. L. Magnanti, and C. L. M. and G. L. Nemhauser (Eds.), *Handbooks in Operations Research and Management Science*, Volume 7, Chapter 6, pp. 401–424. Amsterdam: North-Holland.
- Soboroff, I. M., C. Nicholas, and J. M. Kukla (1997, November). Visualizing document authorship using N-grams and latent semantic indexing. See Ebert and Nicholas (1997), pp. 43–48. ISBN 1581130511. <http://www.cs.umbc.edu/~ian/pubs/hlsi.ps.gz>.
- Sommer, O., A. Dietz, R. Westermann, and T. Ertl (1998, February). An interactive visualization and navigation tool for medical volume data. In V. Skala (Ed.), *The Sixth International Conference in Central Europe on Computer Graphics and Visualization*, Volume 2, pp. 362–371. ResearchIndex: <http://citeseer.nj.nec.com/sommer98interactive.html>.
- Somorjai, R. L., M. Jarmasz, R. Baumgartner, and W. Richter (1999, June). Exploratory analysis of fMRI images: Voxel presentation via “self-similarity”. See Rosen, Seitz, and Volkmann (1999), pp. S46. ISSN 1053–8119.
- Somorjai, R. L., R. Vivanco, N. Pizzi, and M. Jarmasz (2001, June). Direct spatio-temporal analysis of fMRI experiments — a new approach. *NeuroImage* **13**(6, part 2): S252. <http://www.apnet.com/www/journal/hbm2001/10530.html>.
- Sonka, M., V. Hlavac, and R. Boyle (1993). *Image, Processing, Analysis and Machine Vision*. Chapman & Hall.
- Sparck Jones, K. (1972, March). A statistical interpretation of term specificity and its application in retrieval. *Journal of Documentation* **28**(1): 11–21.
- Specht, D. F. (1990). Probabilistic neural networks. *Neural Networks* **3**(1): 109–118.

- Spelke, E. and S. Dehaene (1999, October). Biological foundations of numerical thinking. response to t. j. simon (1999). *Trends in Cognitive Sciences* **3**(10): 365–366.
- Spiegel, M. R. (1968). *Mathematical Handbook of Formulas and Tables*. Schaum's outline series. New York: McGraw-Hill publishing company.
- Stamatakis, E. A., J. T. L. Wilson, and D. J. Wyper (2000, May). Nonlinear spatial normalization of SPECT images with SPM'99. See Fox and Lancaster (2000), pp. S527. ISSN 1053-8119. <http://www.academicpress.com/www/journal/hbm2000/6494.html>.
- Steenblik, R. A. (1986, July). Stereoscopic process and apparatus. U.S. Patent No. 4,597,634.
- Stephan, K. E., L. Kamper, A. Bozkurt, G. A. P. C. Burns, M. P. Young, and R. Kötter (2001, August). Advanced database methodology for the collation of connectivity data on the macaque brain (CoCoMac). *Philosophical Transactions for the Royal Society, London, Series B, Biological Sciences* **356**(1412): 1159–1186. PMID: 11545697. http://www.cocomac.org/cocomac_paper.pdf.
- Stetter, M., I. Schiessl, T. Otto, F. Singpiel, M. Hübener, T. Bonhoeffer, and K. Obermayer (2000, May). Principal component analysis and blind separation of sources for optical imaging of intrinsic signals. *NeuroImage* **11**(5): 482–490. PMID: 10806034. <http://www.idealibrary.com/links/doi/10.1006/nimg.2000.0551/>. ResearchIndex: <http://citeseer.nj.nec.com/stetter00principal.html>.
- Stone, J. V., J. Porrill, C. Büchel, and K. J. Friston (1999a, June). Spatial and temporal independent component analysis of fMRI data. See Rosen, Seitz, and Volkmann (1999), pp. S92. ISSN 1053-8119.
- Stone, J. V., J. Porrill, C. Büchel, and K. J. Friston (1999b, July). Spatial, temporal, spatiotemporal independent component analysis of fMRI data. In *Spatial-temporal Modelling and its applications. 18th Leeds Annual Statistics Research Workshop*. Department of Statistics, University of Leeds. http://www.shef.ac.uk/~pc1jvs/abstracts/ica_porrill_leeds99.html.
- Stone, J. V., J. Porrill, N. R. Porter, and N. M. Hunkin (2000). Spatiotemporal ICA of fMRI data. Computational Neuroscience Report 202, Psychology Department, Sheffield University, Sheffield, UK. ftp://ftp.shef.ac.uk/pub/misc/personal/pc1jvs/papers/stica_nips2000.ps.gz.
- Stone, M. (1974). Cross-validation choice and assessment of statistical predictions. *Journal of the Royal Statistical Society, Series B* **36**: 111–147.
- Strother, S. C., J. R. Anderson, K. A. Schaper, J. J. Sidtis, J. S. Liow, R. P. Woods, and D. A. Rottenberg (1995a, September). Principal component analysis and the scaled subprofile model compared to intersubject averaging and statistical parametric mapping: I. “functional connectivity” of the human motor system studied with [¹⁵O]water PET. *Journal of Cerebral Blood Flow and Metabolism* **15**(5): 738–753.
- Strother, S. C., J. R. Anderson, X. L. Xu, J. S. Liow, D. C. Bonar, and D. A. Rottenberg (1994, November-December). Quantitative comparisons of image registration techniques based on high-resolution MRI of the brain. *Journal of Computer Assisted Tomography* **18**(6): 954–962. PMID: 7962808.
- Strother, S. C., I. Kanno, and D. A. Rottenberg (1995b, May). Commentary and opinion: I. principal component analysis, variance partitioning, and “functional connectivity”. *Journal of Cerebral Blood Flow and Metabolism* **15**(3): 353–360. PMID: 7713992.
- Strother, S. C., N. Lange, J. R. Anderson, K. Rehm, L. K. Hansen, and D. A. Rottenberg (1997a). Activation pattern reproducibility: Measuring the effects of group size and data analysis models. *Human Brain Mapping* **5**(4): 312–316. <http://www3.interscience.wiley.com/cgi-bin/abstract/56436/START>.
- Strother, S. C., N. Lange, J. R. Anderson, K. A. Schaper, and D. A. Rottenberg (1997b, May). Measuring activation pattern reproducibility using resampling techniques. See Friberg, Gjedde, Holm, Lassen, and Nowak (1997), pp. S465. ISSN 1053-8119.
- Strother, S. C., N. Lange, R. Savoy, J. R. Anderson, J. J. Sidtis, L. K. Hansen, P. A. Bandettini, K. O'Craven, M. Rezza, B. R. Rosen, and D. A. Rottenberg (1996a, June). Multidimensional state-spaces for fMRI and PET activation studies. See Belliveau, Fox, Kennedy, Rosen, and Ungerleider (1996), pp. S98. ISSN 1053-8119. http://www.pet.med.va.gov:8080/papers/hbm96_ss.html, <http://fim.nimh.nih.gov/A51.pdf>.
- Strother, S. C., J. J. Sidtis, J. R. Anderson, L. K. Hansen, K. A. Schaper, and D. A. Rottenberg (1996b). [¹⁵O]water PET: More “noise” than signal? In T. Jones, V. Cunningham, R. Myers, and D. Bailey (Eds.), *Quantification of Brain Function Using PET*, San Diego. Academic Press.

- Strupp, J. P. (1996, June). Stimulate: A GUI based fMRI analysis software package. See Belliveau, Fox, Kennedy, Rosen, and Ungeleider (1996), pp. S607. ISSN 1053-8119.
- Svarer, C., L. K. Hansen, and J. Larsen (1993). On the design and evaluation of tapped-delay lines neural networks. In *Proceedings of the IEEE International Conference on Neural Networks, San Francisco, California, USA*, Volume 1, pp. 46-51. [ftp://eivind.imm.dtu.dk](http://eivind.imm.dtu.dk).
- Switzer, P. and A. A. Green (1984). Min/max autocorrelation factors for multivariate spatial imagery. Technical Report 6, Department of Statistics, Stanford University, Stanford, California.
- Talairach, J. and G. Szikla (1967). *Atlas d'Anatomie Stereotaxique du Telencephale: Etudes Anatomo-Radiologiques*. Paris: Masson & Cie.
- Talairach, J. and P. Tournoux (1988, January). *Co-planar Stereotaxic Atlas of the Human Brain*. New York: Thieme Medical Publisher Inc. ISBN 0865772932.
- Tarassenko, L., P. Hayton, N. Cerneaz, and M. Brady (1995). Novelty detection for the identification of masses in mammograms. In *Proceedings of the Fourth International IEE Conference on Artificial Neural Networks*, Volume 409, pp. 442-447. IEE Conference Publication.
- Ter-Pogossian, M. M., M. E. Phelps, E. J. Hoffman, and N. A. Mullani (1975). A positron emission tomograph for nuclear imaging (pett). *Radiology* **114**: 89-98.
- Tesauro, G., D. Touretzky, and T. Leen (Eds.) (1995). *Advances in Neural Information Processing Systems: Proceedings of the 1994 Conference*, Cambridge, Massachusetts. MIT Press.
- The International Society for Cerebral Blood Flow and Metabolism (1999). *Journal of Cerebral Blood Flow and Metabolism*, Volume 19 (supplement 1), Philadelphia. The International Society for Cerebral Blood Flow and Metabolism: Lippincott Williams & Wilkins. Brain '99 and BrainPET '99, XIXth Interation Symposium on Cerebral Blood Flow, Metabolism and Function, IVth Interation Conference on Quantification of Brain Functional with PET.
- Thierry, G., K. Boulanouar, F. Kherif, J. P. Ranjeva, and J. F. Demonet (1999, August). Temporal sorting of neural components underlying phonological processing. *NeuroReport* **10**(12): 2599-2603. PMID: 10574376.
- Thinkmap (2001, March). *Thinkmap white paper*. 157 chamber st., New York: Thinkmap. http://www.thinkmap.com/pdf/TM_whitepaper_english.pdf.
- Thorndike, R. L. (1953). "Who belongs in a family?". *Psychometrika* **18**: 267-276.
- Tibshirani, R. J. (1992). Principal curves revisited. *Statistics and Computing* **2**: 183-190.
- Tibshirani, R. J. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B* **58**(1): 267-288.
- Tikhonov, A. N. (1963). Solution of incorrectly formulated problems and the regularization method. *Soviet Math. Dokl.* **4**: 1035-1038.
- Tikhonov, A. N. and V. Y. Arsenin (1977). *Solution of Ill-Posed Problems*. Washington D. C.: Winston.
- Tipping, M. E. and C. M. Bishop (1997, September). Probabilistic principal component analysis. Technical Report NCRG/97/010, Neural Computing Research Group, Aston University, Aston St, Birmingham, B4 7ET, UK.
- Toft, P. (1996, June). *The Radon Transform — Theory and Implementation*. Ph. D. thesis, Electronics Institute, Technical University of Denmark, Lyngby, Denmark. <http://www.sslug.dk/~pto/cv.html>.
- Toft, P., L. K. Hansen, F. Å. Nielsen, S. C. Strother, N. Lange, N. J. S. Mørch, C. Svarer, O. B. Paulson, R. Savoy, B. R. Rosen, E. Rostrup, and P. Born (1997, May). On clustering of fMRI time series. See Friberg, Gjedde, Holm, Lassen, and Nowak (1997), pp. S456. ISSN 1053-8119. <http://eivind.imm.dtu.dk>.
- Toft, P., F. Å. Nielsen, M. G. Liptrot, and L. K. Hansen (2001, March). *Lyngby fMRI analysis Matlab toolbox* (Version 1.04 ed.). Lyngby, Denmark: Informatics and Mathematic Modelling, Technical University of Denmark. <http://hendrix.imm.dtu.dk/software/lyngby/>.
- Touretzky, D. S., M. C. Mozer, and M. E. Hasselmo (Eds.) (1996). Cambridge, Massachusetts. MIT Press. NIPS-8.
- Tourtzky, D., G. E. Hinton, and T. J. Sejnowski (Eds.) (1988). *Proceedings of the 1988 Connectionist Models Summer School*, San Mateo, CA. Morgan Kauffmann.

- Towle, V. L., I. Syed, C. Berger, R. Grzesczczuk, J. Milton, R. K. Erickson, P. Cogen, E. Berkson, and J.-P. Spire (1998, January). Identification of the sensory/motor area and pathologic regions using ECoG coherence. *Electroencephalogr Clin Neurophysiol* **106**(1): 30–39.
- Toyama, H., K. Takazawa, T. Nariel, K. Uemura, and M. Senda (1999). Visualization of correlated hemodynamic and metabolic functions in cerebrovascular disease by a cluster analysis with PET study. See The International Society for Cerebral Blood Flow and Metabolism (1999), pp. S810. Brain '99 and BrainPET '99, XIXth Interation Symposium on Cerebral Blood Flow, Metabolism and Function, IVth Interation Conference on Quantification of Brain Functional with PET.
- Toyama, H., K. Takazawa, T. Nariel, K. Uemura, and M. Senda (2001). Visualization of correlated hemodynamic and metabolic functions in cerebrovascular disease by a cluster analysis with PET study. Chapter 44, pp. 301–304.
- Tuch, D. S., J. W. Belliveau, and V. J. Wedeen (2000, May). Probabilistic tractography using high angular resolution diffusion imaging. See Fox and Lancaster (2000), pp. S913. ISSN 1053-8119.
- Tulving, E., S. Kapur, F. I. M. Craik, M. Moscovitch, and S. Huole (1994a). Hemispheric encoding/retrieval asymmetry in episodic memory: positron emission tomography findings. *Proceedings of the National Academy of Sciences* **91**: 2016–2020.
- Tulving, E., S. Kapur, H. J. Markowitsch, F. I. M. Craik, R. Habib, and S. Houle (1994b, March). Neuroanatomical correlates of retrieval in episodic memory: Auditory sentence recognition. *Proceedings of the National Academy of Sciences of the USA* **91**: 2012–2015.
- Turkeltaub, P., G. F. Eden, K. M. Jones, and T. A. Zeffiro (2001, June). A novel meta-analysis technique applied to single word reading. *NeuroImage* **13**(6 (part 2)): S272. <http://www.apnet.com/www/journal/hbm2001/10478.html>.
- Turkheimer, F., K. Pettigrew, L. Sokoloff, C. B. Smith, and K. Schmidt (2000, August). Selection of an adaptive test statistic for use with multiple comparison analyses of neuroimaging data. *NeuroImage* **12**(2): 219–229.
- Turkington, T. G., R. J. Jazczak, C. A. Pelizzari, C. C. Harris, J. R. MacFall, J. M. Hoffman, and R. E. Coleman (1993, September). Accuracy of registration of PET, SPECT and MR images of a brain phantom. *Journal of Nuclear Medicine* **34**(9): 1587–1594. PMID: 8355080.
- Turner, R., D. Le Bihan, C. T. W. Moonen, D. Despres, and J. Frank (1991, November). Echo-planar time course mri of cat brain deoxygenation changes. *Magnetic Resonance in Medicine* **22**(1): 159–166. PMID: 1798390.
- Van den Wollenberg, A. L. (1977). Redundancy analysis: An alternative for canonical correlation analysis. *Psychometrika* **42**: 207–219.
- Van Essen, D. C. and H. A. Drury (1997, September). Structural and functional analyses of human cerebral cortex using a surface-based atlas. *Journal of Neuroscience* **17**(18): 7079–7102. PMID: 9278543. <http://www.jneurosci.org/cgi/content/full/17/18/7079>.
- Van Essen, D. C., H. A. Drury, and C. H. Anderson (1999, June). An automated method for accurately reconstructing the cortical surface. See Rosen, Seitz, and Volkmann (1999), pp. S173. ISSN 1053–8119.
- Van Gelder, A. and K. Kim (1996, October). Direct volume rendering with shading via three-dimensional textures. In *1996 Volume Visualization Symposim*, pp. 23–300. IEEE. ISBN 0-89791-741-3. <http://www.cse.ucsc.edu/~ksk/resumes/voltx96.ps>. ResearchIndex: <http://citeseer.nj.nec.com/334285.html>.
- Vanzetta, I. and A. Grinvald (1999, November). Increased cortical oxidative metabolism due to sensory stimulation: implications for functional brain imaging. *Science* **286**(5444): 1555–1558. PMID: 10567261.
- Vanzetta, I. and A. Grinvald (2001, June). Evidence and lack of evidence for the initial dip in the anesthetized rat: Implications for human functional brain imaging. *NeuroImage* **13**(6): 959–967.
- Vasileios Megalooikonomou, C. D. and E. H. Herskovits (1999, August). Mining lesion-deficit associations in a brain image database. In *Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining, August 15 - 18, 1999, San Diego, CA USA*, pp. 347–351. ResearchIndex: <http://citeseer.nj.nec.com/333930.html>.
- Vazquez, A. L. and D. C. Noll (1998, February). Nonlinear aspects of the bold response in functional mri. *NeuroImage* **7**(2): 108–118.
- Veltman, D. J., K. J. Friston, G. Sanders, and C. J. Price (2000, June). Regionally specific sensitivity differences in fMRI and PET: Where do they come from? *NeuroImage* **11**(6): 575–588.

- Vérard, L., P. Allain, J. M. Travère, J. C. Baron, and D. Bloyet (1997, October). Fully automatic identification of AC and PC landmarks on brain MRI using scene analysis. *IEEE Transactions on Medical Imaging* **16**(5): 610–616. PMID: 9368116.
- Verwichte, E. and K. Galsgaard (1998). On the visualization of three-dimensional datasets. *Solar Physics* **183**(2): 445–448.
- Villringer, A., B. R. Rosen, J. W. Belliveau, J. L. Ackermann, R. B. Lauffer, R. B. Buxton, Y.-S. Chao, V. J. Wedeen, and T. J. Brady (1988, February). Dynamic imaging of lanthanide chelates in normal brain: Changes in signal intensity due to susceptibility effects. *Magnetic Resonance in Medicine* **6**(2): 164–174. PMID: 3367774.
- Vinod, H. (1969). Integer programming and the theory of grouping. *Journal of the American Statistical Association* **40**: 1098–1099.
- Vinod, H. D. (1976). Canonical ridge and econometrics of joint production. *J. Econometrics* **4**: 147–166.
- Vlontzos, J., J.-N. Hwang, and E. Wilson (Eds.) (1994). Piscataway, New Jersey. IEEE.
- Volbracht, S., G. Domik, K. Shahrabaki, and G. Fels (1997, March). How effective are 3d display modes? In *Conference on Human Factors in Computing Systems — Proceedings*, pp. 540–541. ACM. <http://www.acm.org/sigchi/chi97/proceedings/tech-note/sbv.htm>. Technical note.
- von Economo, C. (1929). *The cytoarchitectonics of the human cerebral cortex*. London: Humphrey Milford; Oxford University Press.
- Wada, J. and T. Rasmussen (1960). Intracarotid injections of sodium amytal for the lateralization of cerebral speech dominance. *J. Neurosurg.* **17**: 266–282.
- Wagner, R. A. and M. J. Fischer (1974, January). The string-to-string correction problem. *Journal of the Association for Computing Machinery* **21**(1): 168–173.
- Walter, H., R. Kristeva, U. Knorr, G. Schlaug, Y. Huang, H. Steinmetz, B. Nebeling, H. Herzog, and R. J. Seitz (1992, Winter). Individual somatotopy of primary sensorimotor cortex revealed by intermodal matching of MEG, PET, and MRI. *Brain Topography* **5**(2): 183–187. PMID: 1489648.
- Ward, B. D. (1997, June). *Simultaneous Inference for FMRI Data*. Biophysics Research Institute, Medical College of Wisconsin. <http://afni.nimh.nih.gov/afni/docps/AlphaSim.ps>.
- Watabe, H. (2001, May). Gpetview. Internet. <http://homepage2.nifty.com/peco/gpetview/gpetview.html>.
- Weaver, J. B. (1995, September). Efficient calculation of the principal components of imaging data. *J Cereb Blood Flow Metab* **15**(5): 892–894. PMID: 7673383.
- Weiller, C., C. Isensee, M. Rijntjes, W. Huber, S. Muller, D. Bier, K. Dutschka, R. P. Woods, J. Noth, and H. C. Diener (1995, June). Recovery from Wernicke's aphasia: a positron emission tomographic study. *Annals of Neurology* **37**(6): 723–732. PMID: 7778845.
- Weiss, R., B. Velez, M. A. Sheldon, C. Nemprenpre, P. Szilagyi, A. Duda, and D. K. Gifford (1996). HyPursuit: A hierarchical network search engine that exploits content-link hypertext clustering. In *Proceedings of the Seventh ACM Conference on Hypertext, Washington, DC, March 1996*. <http://www.psrg.lcs.mit.edu/Images/PDFIcon.gif>.
- Wennerås, C. and A. Wold (1997). Nepotism and sexism in peer-review. *Nature* **387**: 341–343.
- Wentz, K. U., H. P. Mattle, R. R. Edelman, J. Kleefield, G. V. O'Reilly, C. Liu, and B. Zhao (1991). Stereoscopic display of MR angiograms. *Neuroradiology* **33**(2): 123–125. PMID: 2046895.
- Wessinger, C. M., J. VanMeter, B. Tian, J. Van Lare, J. Pekar, and J. P. Rauschecker (2001, January). Hierarchical organization of human auditory cortex revealed by functional magnetic resonance imaging. *Journal of Cognitive Neuroscience* **13**(1): 1–7.
- West, J., J. M. Fitzpatrick, M. Y. Wang, B. M. Dawant, C. R. Maurer Jr, R. M. Kessler, R. J. Maciunas, C. Barillot, D. Lemoine, A. Collignon, F. Maes, P. Suetens, D. Vandermeulen, P. A. van den Elsen, S. Napel, T. S. Sumanaweera, B. Harkness, P. F. Hemler, D. L. Hill, D. J. Hawkes, C. Studholme, J. B. Maintz, M. A. Viergever, G. Malandain, and R. P. Woods (1997, July-August). Comparison and evaluation of retrospective intermodality brain image registration techniques. *Journal of Computer Assisted Tomography* **21**(4): 554–566. PMID: 9216759. http://www.vuse.vanderbilt.edu/~image/registration/online_files/jcat_97.pdf.

- Wexelblat, A. (1998). History-rich tools for social navigation. In *CHI'98*. ACM Press. <http://wex.www.media.mit.edu/people/wex/Footprints2/fp-v2.html>. Summary.
- White, H. D. (1981). Cocited author retrieval online: An experiment with the social indicators literature. *Journal of the American Society for Information Science* **32**: 16–22.
- White, H. D. and B. C. Griffith (1981). Author cocitation: A literature measure of intellectual structure. *Journal of the American Society for Information Science* **32**: 163–171.
- White, H. D. and K. W. McCain (1998, April). Visualizing a discipline: an author co-citation analysis of information science, 1972–1995. *Journal of the American Society for Information Science* **49**(4): 327–355.
- Widman, G., C. Helmstaedter, K. Lehnertz, and C. E. Elgar (1999, June). Nonlinear ECoG analysis indicates focal and specific focal and specific changes in cortical activation during ongoing behaviour. See Rosen, Seitz, and Volkmann (1999), pp. S111. ISSN 1053–8119.
- Wiegell, M. R. (2000, May). *Diffusion Tensor MRI of Brain Anatomy in Health and Disease*. Ph. D. thesis, Faculty of Health Sciences, University of Copenhagen, Copenhagen, Denmark. http://www.nmr.mgh.harvard.edu/lbsf/publications/Thesis_MRWiegell.pdf.gz.
- Wiegell, M. R., D. S. Tuch, H. B. W. Larsson, and V. J. Wedeen (2000, May). Automatic identification of thalamic nuclei from DTI. See Fox and Lancaster (2000), pp. S453. ISSN 1053–8119. <http://www.apnet.com/www/journal/hbm2000/6490.html>.
- Wilbur, W. J. and Y. Yang (1996, May). An analysis of statistical term strength and its use in the indexing and retrieval of molecular biology texts. *Comput. Biol. Med.* **26**(3): 209–222. PMID: 8725772.
- Williams, C. K. I. and C. E. Rasmussen (1996). Gaussian processes for regression. See Touretzky, Mozer, and Hasselmo (1996). ftp://ftp.cs.toronto.edu/pub/carl/gauss_process_nips.ps.gz. NIPS-8.
- Williams, P. M. (1995). Bayesian regularization and pruning using a Laplace prior. *Neural Computation* **7**: 117–143.
- Wills, A. J., I. H. Jenkins, P. D. Thompson, L. J. Findley, and D. J. Brooks (1994, October). Red nuclear and cerebellar but no olivary activation associated with essential tremor: a positron emission tomographic study. *Annals of Neurology* **36**(4): 636–642. ISSN 0364–5134. PMID: 7944296.
- Wilson, O., A. Van Gelder, and J. Wilhelms (1994). Direct volume rendering via 3D textures. Technical Report UCSC-CRL-94-19, University of California, Santa Cruz.
- Wise, R., F. Chollet, U. Hadar, K. J. Friston, E. Hoffner, and R. S. J. Frackowiak (1991). Distribution of cortical neural networks involved in word comprehension and word retrieval. *Brain* **114**(Pt 4): 1803–1817. PMID: 1884179.
- Wold, H. (1966). Estimation of principal components and related models by iterative least squares. In P. Krishnaiah (Ed.), *Multivariate Analysis*, pp. 391–420. New York: Academic Press.
- Wold, H. (1975). Soft modeling by latent variables, the nonlinear iterative partial least squares approach. In J. Cani (Ed.), *Perspectives in Probability and Statistics, Papers in Honour of M. S. Bartlett*. Academic Press.
- Wold, S. (1978). Cross-validatory estimation of the number of components in factor and principal component models. *Tecnometrics* **20**(4): 397–405.
- Wold, S., A. Ruhe, H. Wold, and W. J. Dunn III (1984, September). The collinearity problem in linear regression. the partial least squares (PLS) approach to generalized inverses. *SIAM J. Sci. Stat. Comput.* **5**(3): 735–743.
- Wolfe, P. (1968). Another variable metric method. Working paper.
- Wolpert, D. H. and W. G. Macready (1997). No free lunch theorems for optimization. *IEEE Transactions on Evolutionary Computation* **1**(1): 67–82. ResearchIndex: <http://citeseer.nj.nec.com/wolpert96no.html>.
- Woods, R. P., S. R. Cherry, and J. C. Mazziotta (1992, July-August). Rapid automated algorithm for alignment and reslicing PET images. *Journal of Computer Assisted Tomography* **16**(4): 634–639. PMID: 1629424.
- Woods, R. P., M. Dapretto, N. L. Sicotte, A. W. Toga, and J. C. Mazziotta (1999). Creation and use of a Talairach-compatible atlas for accurate, automated, nonlinear intersubject registration, and analysis of functional imaging data. *Human Brain Mapping* **8**(2–3): 73–79. ISSN 1097–0193. PMID: 10524595. <http://www3.interscience.wiley.com/cgi-bin/abstract/66000625/>. Special Issue: Proceedings of the Brainmap '98 Workshop.

- Woods, R. P., S. T. Grafton, C. J. Holmes, S. R. Cherry, and J. C. Mazziotta (1998a). Automated image registration. I. general methods and intrasubject, intramodality validation. *Journal of Computer Assisted Tomography* **22**(1): 139–152. PMID: 9448779.
- Woods, R. P., S. T. Grafton, J. D. G. Watson, N. L. Sicotte, and J. C. Mazziotta (1998b). Automated image registration. II. intersubject validation of linear and nonlinear models. *Journal of Computer Assisted Tomography* **22**(1): 153–165. PMID: 9448780.
- Woods, R. P., J. C. Mazziotta, and S. R. Cherry (1993). MRI-PET registration with automated algorithm. *Journal of Computer Assisted Tomography* **17**(4): 536–546.
- Worden, K. (1997). Structural fault detection using a novelty measure. *Journal of Sound and Vibration* **201**(1): 85–101.
- Wormell, I. (1998, December). Informetric analysis of the international impact of scientific journals: How “international” are the international journals? *Journal of Documentation* **54**(5): 584–605.
- Worsley, K. J. (1994). Local maxima and the expected Euler characteristic of excursion sets of χ^2 , f and t fields. *Advances in Applied Probability* **26**: 13–42.
- Worsley, K. J. (1995). Boundary corrections for the expected Euler characteristic of excursion sets of random fields, with an application to astrophysics. *Advances in Applied Probability* **27**: 943–959.
- Worsley, K. J. (1997). An overview and some new developments in the statistical analysis of PET and fMRI data. *Human Brain Mapping* **5**(4): 254–258.
- Worsley, K. J., M. Andermann, T. Koulis, D. MacDonald, and A. C. Evans (1999). Detecting changes in non-isotropic images. *Human Brain Mapping* **8**: 98–101.
- Worsley, K. J., A. C. Evans, S. Marrett, and P. Neelin (1992, November). A three-dimensional statistical analysis for CBF activation studies in human brain. *Journal of Cerebral Blood Flow and Metabolism* **12**(6): 900–918. PMID: 1400644. <http://euclid.math.mcgill.ca/~keith/jcbf/jcbf.ps.gz>.
- Worsley, K. J. and K. J. Friston (1995, September). Analysis of fMRI time-series revisited — again. *NeuroImage* **2**(3): 173–181. PMID: 9343600. http://www.fil.ion.ucl.ac.uk/spm/papers/fMRI_4, <http://www.idealibrary.com/links/doi/10.1006/nimg.1995.1023/pdf>.
- Worsley, K. J., C. Liao, M. Grabove, V. Petre, B. Ha, and A. C. Evans (2000, June). A general statistical analysis for fMRI data. *NeuroImage* **11**(5, part 2): S648. <http://www.math.mcgill.ca/keith/HBM2000/HBM2000.html>.
- Worsley, K. J., S. Marrett, P. Neelin, and A. C. Evans (1996). Searching scale space for activation in PET. *Human Brain Mapping* **4**: 74–90. <http://www.math.mcgill.ca/~keith/scale/scale.abstract.html>.
- Worsley, K. J., J.-B. Poline, K. J. Friston, and A. C. Evans (1997, November). Characterizing the response of PET and fMRI data using multivariate linear models. *NeuroImage* **6**(4): 305–319. PMID: 9417973. <http://www.idealibrary.com/links/citation/1053-8119/6/305>.
- Xiong, J., J.-H. Gao, J. L. Lancaster, and P. T. Fox (1995). Clustered pixels analysis for functional MRI activation studies of the human brain. *Human Brain Mapping* **3**: 287–301.
- Xiong, J., J.-H. Gao, J. L. Lancaster, and P. T. Fox (1996). Assessment and optimization of functional MRI analyses. *Human Brain Mapping* **4**(3): 153–167.
- Xiong, J., S. Rao, J. H. Gao, M. Woldorff, and P. T. Fox (1998). Evaluation of hemispheric dominance for language using functional MRI: a comparison with positron emission tomography. *Human Brain Mapping* **6**(1): 42–58. PMID: 9673662.
- Yacoub, E. and X. Hu (1999, June). Detection of the early negative response in fMRI at 1.5 Tesla. *Magnetic Resonance in Medicine* **41**(6): 1088–1092. PMID: 10371439. <http://www3.interscience.wiley.com/cgi-bin/abstract/62001111/START>.
- Yacoub, E. and X. Hu (2001, February). Detection of the early decrease in fmri signal in the motor area. *Magnetic Resonance in Medicine* **45**(2): 184–190. ISSN 0740-3194. PMID: 11180423.
- Yacoub, E., T. H. Le, K. Ugurbil, and X. Hu (1999). Further evaluation of the initial negative response in functional magnetic resonance imaging. *Magnetic Resonance in Medicine* **41**: 436–441.
- Yee, T. W. and T. J. Hastie (2000, August). Reduced-rank vector generalized linear models. Internet. <http://www-stat.stanford.edu/~hastie/Papers/rrr.ps>. ResearchIndex: <http://citeseer.nj.nec.com/323845.html>.

- Young, M. P. (1992, July). Objective analysis of the topological organization of the primate cortical visual system. *Nature* **358**(6382): 152–155. PMID: 1614547.
- Young, P. (1996). Three dimensional information visualisation. Computer Science Technical Report 12/96, Visualisation Research Group, Centre for Software Maintenance, Department of Computer Science, University of Durham.
- Ypma, A. and R. P. W. Duin (1998). Novelty detection using self-organizing maps. In N. Kasabov, R. Kozma, K. Ko, R. O'Shea, G. Coghill, and T. Gedeon (Eds.), *Progress in Connectionist-Based Information Systems. Proceedings of the 1997, International Conference on Neural Information Processing and Intelligent Information Systems (ICONIP97)*, Volume 2, Singapore, pp. 1322–1325. IEE: Springer Verlag. ISBN 9833083646. http://www.ph.tn.tudelft.nl/~ypma/papers/list_of_papers.html.
- Zarahn, E., G. K. Aguirre, and M. D'Esposito (1997, April). Empirical analyses of BOLD fMRI statistics. *NeuroImage* **5**(3): 179–197.
- Zatorre, R. J., M. Jones-Gotman, A. C. Evans, and E. Meyer (1992, November). Functional localization and lateralization of human olfactory cortex. *Nature* **360**(6402): 339–340. PMID: 1448149.
- Zavaljevski, A., A. P. Dhawan, M. Gaskil, W. Ball, and J. D. Johnson (2000, March-April). Multi-level adaptive segmentation of multi-parameter MR brain images. *Comput. Med. Imaging. Graph.* **24**(2): 87–98. PMID: 10767588.
- Zelevnik, M. P. (1997, May). 3d visualization: What does it mean? In *The XII International Conference on the Use of Computers in Radiation Therapy (XII ICCR)*, May 27-30, 1997, Salt Lake City, UT. <http://rtp2.med.utah.edu/~zelevnik/vispaper/paper.html>.
- Zhang, Y., M. Brady, and S. Smith (2001). Segmentation of brain MR images through a hidden Markov random field model and the expectation maximization algorithm. *IEEE Transaction of Medical Imaging* **20**(1): 45–57. <http://www.fmrib.ox.ac.uk/analysis/techrep/tr00yz1/tr00yz1.ps.gz>.
- Zhu, W., N. Wolkow, K. Mueller, T. Welsh, and J. Meade (2001, June). BrainMiner — an interactive tool for functional connectivity study. *NeuroImage* **13**(6): S293. <http://www.apnet.com/www/journal/hbm2001/10525.html>.
- Zilles, K. and N. Palomero-Gallagher (2001, July). Cyto-, myelo-, and receptor architectonics of the human parietal cortex. *NeuroImage* **14**(1, part 2): S8–S20. PMID: 11373127. <http://www.idealibrary.com/links/citation/1053-8119/14/S8>.
- Zipf, G. K. (1949). *Human Behavior and the Principle of Least Effort*. Reading, Massachusetts: Addison-Wesley.

Author Index

- Abato, Michael 94, 283
Abbott, David 93, 267
Abel, T. 25, 285
Abernathey, Chad D. 23, 303
Ackermann, Jerome L. 30, 315
Adams, K. 18, 291
Aguirre, Geoffrey Karl 25, 45, 46, 48, 90, 98, 103, 108, 267, 279, 318
Ahmed, Ayesha 24, 278
Aho, Alfred V. 100, 267
Akaho, Shotaro 51, 267
Akaike, H. 40, 267
Akaike, Hirotugu 40, 267
Akbulak, Erbil 22, 285
Akeley, Kurt 85, 267
Åkerman, S. 41, 296
Albert, N. M. 125, 292
Albright, Kevin 31, 283
Alexander, M. 46, 309
Allain, Pascal 114, 315
Allison, Truett 27, 28, 98, 267
Allman, J. M. 81, 280
Almeida, Rita 52, 267
Almind, Tomas C. 96, 267
Alperovitch, Annick 44, 306
Alpert, N. M. 22, 307
Alpert, Nathaniel M. 123, 292
Altschul, S. F. 109, 267
Amari, Shun-ichi 36, 40, 56, 267, 268, 301
Amunts, K. 44, 94, 96, 281, 308
Amunts, Katrin 23, 24, 268, 301, 308
Andermann, M. 41, 317
Andersen, Anders H. 55, 268
Andersen, Lars Nonboe 35, 38, 63, 268, 294
Andersen, Peter M. 28, 307
Anderson, Adam K. 123, 292
Anderson, C. H. 45, 314
Anderson, Charles H. 45, 278
Anderson, Jon R. 39, 42–45, 49, 54, 55, 71, 78–80, 90, 91, 117, 157, 200–221, 282, 291, 294, 301, 307, 312
Anderson, P. 26, 300
Anderson, Theodore Wilbur 54, 71, 268
Andersson, Jesper L. R. 43, 46–48, 268
Andersson, M. 44, 308
Andreasen, Nancy C. 48, 93, 268, 284
Ardekani, Babak 43, 268
Ardekani, Babak A. 46, 47, 268, 291
Armitage, P. 41, 69, 268
Armstrong, J. Scott 101, 120, 268
Arndt, Stephan 48, 93, 268, 284
Arnfred, Sidse 28, 268
Arnold, J. B. 42, 269, 309
Arnold, James B. 78–80, 282
Arsenin, V. Y. 37, 313
Ashburner, John 23, 28, 38, 42–46, 48, 52, 125, 268, 269, 281, 285, 291, 296, 298, 299, 306
Askew, S. 57, 309
Aslam, Javed A. 94, 96, 285
Asoh, Hideki 51, 79, 267, 269
Aston, J. 52, 285
Aston, J. A. D. 93, 269
Athwal, B. S. 25, 60, 69, 299
Atkinson, Jeff 25, 288
Atkinson, Richard C. 125, 269
Atkinson, Rita L. 125, 269
Attias, Hagai 56, 269
Avis, David 45, 297
Avison, M. 25, 306
Avison, Malcolm J. 55, 268

Bailey, Dale 42, 299
Bakay, Roy. A. E. 18, 291
Baker, P. 26, 289
Baldi, P. 74, 269
Ball, W. 45, 318
Balslev, Daniela 51, 52, 157, 269
Baluja, Shumeet 70, 269
Banati, R. B. 52, 301
Bancaud, J. 109, 294
Bandettini, Peter A. 25, 26, 30, 46, 54, 67, 68, 79, 108, 269–271, 312
Banfield, Jeffrey D. 51, 269
Bannister, Peter R. 93, 311
Bañuelos, Verónica Medina 84, 93, 298
Barcys, C. 55, 299
Bargel, B. 95, 303
Barillot, C. 43, 45, 284, 315
Barinaga, Marcia 25, 269
Barker, A. T. 31, 269
Barlow, H. B. 120, 269
Baron, J. C. 114, 315
Bartylla, K. 28, 285
Baumgart, F. 95, 303
Baumgartner, Richard 48, 52, 269, 279, 311
Baune, Axel 52, 269
Bayer, Alan E. 105, 269
Beauchamp, N. J. 52, 297
Becker, J. Alex 94, 290
Becker, James T. 123, 270
Becker, Sue 36, 270
Becker, Suzanna 79, 137, 270
Beckman, R. J. 62, 63, 270
Beckmann, Christian 93, 311
Beckmann, Christian F. 46, 270
Belger, Aysenil 27, 28, 267
Bell, A. J. 57, 299
Bell, Antony J. 57, 270
Bell, Gavin 84, 270, 274

- Bellgowan, P. S. 22, 271
 Bellgowan, Patrick S. F. 68, 270
 Belliveau, J. W. 30, 270, 293, 315
 Belliveau, John W. 30, 314
 Beltrami, E. 53, 270
 Bem, Daryl J. 125, 269
 Ben-Israel, Adi 136, 270
 Benali, Habib 66, 93, 118, 275, 298
 Bengio, Yoshua 38, 50, 270, 272
 Benjamini, Yoav 41, 270
 Berger, C. 28, 314
 Berger, Hans 27, 270
 Berger, James O. 139, 270
 Berks, George 90, 270
 Berkson, E. 28, 314
 Berners-Lee, Tim 18, 270
 Bero, Lisa A. 120, 284
 Berry, Geoffrey 41, 69, 268
 Berry, I. 33, 272
 Bettinardi, V. 109, 305
 Bezdek, J. C. 50, 270
 Bhansali, B. J. 40, 270
 Bibby, John M. 34, 35, 37, 49–51, 53, 54, 63, 64, 66, 70–75, 78, 86, 103, 104, 133–135, 137, 138, 145–148, 151, 153, 298
 Bichteler, Julie 104, 271
 Biederman, J. 143, 274
 Bienenstock, E. 39, 135, 283
 Bienenstock, C. A. 143, 287
 Bier, D. 125, 315
 Biernacki, C. 40, 51, 271
 Biernacki, Christophe 40, 274
 Bilke, Sven 104, 271
 Binder, J. R. 26, 271
 Binder, Jeff R. 22, 271
 Birbaumer, N. 18, 293
 Birmingham, William P. 97, 284
 Birn, Rasmus M. 26, 271
 Birrell, Andrew 96, 271
 Birznies, Gunther 153, 285
 Bishop, Christopher M. 35–37, 39, 53, 54, 61–63, 135, 136, 139, 271, 313
 Biswal, Bharat B. 60, 271
 Bitter, Ingmar 85, 276
 Björnsson, C. H. 101, 137, 271
 Blinkenberg, M. 25, 308
 Bloch, Felix 29, 271
 Blomqvist, H. 44, 310
 Bloyet, D. 51, 279
 Bloyet, Daniel 114, 315
 Bly, Benjamin Martin 96, 271
 Bobholz, J. A. 93, 284
 Bodurka, Jerzy 30, 271
 Bogerts, B. 95, 303
 Bohm, C. 44, 285, 310
 Bohm, Christian 44, 289
 Boissonnat, Jean-Daniel 86, 271
 Bolas, Mark T. 88, 299
 Bollacker, Kurt 96, 106, 271, 284, 295
 Bolton, J. P. 28, 289
 Bonar, D. C. 43, 312
 Bonhoeffer, T. 30, 55–57, 312
 Bookstein, F. L. 72, 79, 80, 299
 Bor, Daniel 24, 278
 Borge, Magnus 52, 71, 73, 74, 79, 80, 271, 281
 Born, A. Peter 25, 308
 Born, Alfred Peter 17, 26, 30, 271, 272
 Born, Peter 52, 54, 65, 68, 157, 159, 286, 303, 313
 Borodin, Allan 105, 272
 Bosch, Volker 93, 297
 Botafogo, Rodrigo A. 105, 272
 Bottou, Léon 50, 272
 Bouchareine, A. 109, 294
 Bouet, Marienette 96, 278
 Boulanouar, K. 33, 68, 272, 313
 Bowden, Douglas M. 23, 87, 94, 143, 272
 Bowers, Philip L. 90, 289
 Bowtell, R. 26, 289
 Box, George E. P. 40, 66, 69, 272
 Boyle, M. B. 31, 285
 Boyle, Roger 81, 311
 Boynton, Geoffrey M. 26, 65, 272
 Bozkurt, Ahmet 94, 95, 312
 Brady, M. 62, 313
 Brady, Michael 45, 318
 Brady, Mike 93, 311
 Brady, T. J. 30, 270, 293, 315
 Brahe, Tycho 18, 272
 Brammer, M. 79, 274
 Brammer, Michael 65, 273
 Brammer, Michael J. 78–80, 273
 Braun, M. 43, 268
 Breen, Gillian 120, 285
 Breiter, H. C. R. 22, 307
 Brent, R. P. 37, 272
 Brett, Matthew 23, 43, 45, 90, 126, 272, 290, 308
 Bretthorst, G. Larry 66, 272
 Bridle, J. S. 139, 272
 Brin, Sergey 105, 272
 Bro-Nielsen, Morten 41, 272
 Bro, Rasmus 78, 272
 Broca, Paul 24, 272
 Brockway, J. P. 17, 273
 Broder, Andrei 104, 273
 Brodie, J. D. 123, 310
 Brodmann, Korbinian 23, 273
 Brody, Carlos 74, 308
 Bronen, Richard A. 123, 310
 Brooks, David J. 123, 316
 Brooks, Rodney J. 74, 273
 Brown, Greg 57, 299
 Brown, Greg B. 57, 299
 Brown, M. Christian 18, 273
 Brown, Marc H. 83, 301
 Brown, P. J. 37, 273
 Brown, W. Douglas 123, 288
 Broyden, C. G. 36, 273
 Bruckbauer, T. 90, 287
 Bryan, R. N. 94, 273
 Bryan, R. Nick 90, 94, 98, 287, 296, 303
 Bub, D. 122, 123, 295
 Buchbinder, B. R. 30, 270
 Büchel, Christian 42, 57, 71, 273, 282, 288, 312
 Buchsbaum, Bradley R. 45, 283
 Buchsbaum, M. S. 143, 287
 Buchsbaum, Monte S. 45, 283
 Buckley, Chris 99, 101, 102, 309
 Buckner, Randy L. 25, 65, 98, 118, 123, 273, 276, 292
 Bullmore, Edward 65, 273
 Bullmore, Edward T. 62, 78–80, 273, 274, 279
 Buntine, Wray 99, 273
 Burchard, Paul 87, 301
 Burchhardt, J. 46, 305

- Burden, Simon 120, 285
 Bürgel, Uli 23, 24, 268
 Burnham, Alison J. 66, 71, 74, 76, 149, 273
 Burns, Gully A. P. C. 94, 95, 273, 312
 Burt, C. 53, 135, 273
 Busatto, G. 79, 274
 Busch, Nils 93, 297
 Bush, George 98, 143, 274
 Bush, Vannevar 18, 274
 Buxton, R. B. 30, 315
 Buxton, Richard B. 25, 26, 66, 71, 274

 Cabeza, Roberto 98, 274
 Cabral, Brain K. 85, 274
 Cai, W. 52, 279
 Cailliau, Robert 18, 270
 Camici, P. G. 86, 308
 Cao, J. 40, 41, 274
 Cao, Jin 41, 274
 Card, Stuart K. 87, 308
 Carey, Rick 84, 274
 Carman, George J. 68, 309
 Carpenter, Mark P. 99, 105, 274
 Carpenter, T. A. 46, 302
 Carr, Leslie 87, 275
 Carrière, Jeromy 87, 107, 274
 Carrozzo, M. 109, 305
 Cartwright, C. 143, 287
 Cavada, C. 44, 94, 96, 281, 308
 Cedefamn, Jonny 52, 79, 80, 281
 Celeux, Gilles 39, 40, 50, 51, 271, 274
 Celsis, P. 33, 272
 Cerneaz, N. 62, 313
 Chabris, Christopher F. 123, 292
 Chakrabarti, Soumen 96, 101, 274
 Chan, E. 94, 293
 Chance, Britton 31, 274
 Chang, H. 107, 275
 Chang, Shih-Fu 96, 311
 Chao, Y.-S. 30, 315
 Charniak, Eugene 100, 274
 Chawla, Dave 71, 282
 Chen, Andrew C. N. 28, 86, 158, 268, 275
 Chen, Anita 98, 287
 Chen, Baoquan 85, 276
 Chen, C. T. 43, 304
 Chen, Chaomei 87, 107, 275
 Chen, G. T. 43, 304
 Chen, Kewei 52, 287
 Chen, Yue 31, 274
 Cheng, H.-M. 30, 293
 Chernoff, H. 86, 275
 Cherry, S. R. 43, 93, 116, 316, 317
 Chesler, D. A. 28, 30, 66, 290, 293, 303
 Chicurel, Mirana 93, 275
 Chittaro, Luca 83, 275
 Chkhenkeli, Sozari 28, 275
 Chollet, F. 33, 125, 272, 316
 Christian, Brad 93, 284
 Christiansen, Torben B. 157, 269
 Christiansen, Torben Bebensee 51, 52, 61, 157, 269, 275
 Christoff, Kalina 45, 272
 Cizadlo, Ted 48, 93, 268, 284
 Clare, S. 26, 289
 Clare, Stuart 46, 93, 275, 311
 Clarke, C. J. S. 28, 289

 Clifton, G. L. 31, 307
 Cline, Harvey E. 85, 275, 297
 Cocosco, Chris A. 90, 94, 275
 Coetzee, Frans 19, 155, 158, 295
 Coetzee, Frans M. 97, 100, 278, 292
 Cogen, P. 28, 314
 Cohen, David 28, 275
 Cohen, J. D. 41, 279
 Cohen, Jonathan D. 99, 275
 Cohen, L. B. 31, 285
 Cohen, M. S. 30, 42, 269, 270, 293, 309
 Cohen, Mark S. 25, 275
 Cohen, Victor W. 29, 307
 Cohn, David 107, 275
 Cointepas, Yann 93, 275
 Coleman, R. E. 43, 314
 Collignon, A. 43, 298, 315
 Collins, D. L. 43, 44, 269
 Collins, D. Louis 45, 94, 95, 125, 275, 276, 279, 284
 Comaniciu, Dorin 97, 276
 Conradsen, Knut 54, 69, 73, 276
 Constable, R. Todd 17, 78, 79, 123, 291, 299, 310
 Cook, R. Dennis 62, 63, 270
 Corbetta, M. 65, 67, 303
 Corbetta, Maurizio 123, 276
 Cordes, D. 46, 276
 Cormack, Allan M. 28, 276
 Cornhill, J. Frederick 85, 310
 Coryell, C. D. 30, 304
 Cottrell, G. W. 79, 138, 277
 Courtney, Susan M. 90, 287
 Covaert, Gérard 50, 51, 274
 Cowen, Colleen 31, 274
 Cowey, A. 44, 94, 96, 281, 308
 Cox, David R. 103, 105, 276
 Cox, R. W. 22, 271
 Cox, Robert William 46, 93, 276
 Craik, F. I. M. 98, 314
 Craik, Fergus I. M. 123, 314
 Crawley, Adrian P. 17, 277
 Crelieu, Gérard R. 25, 288
 Crivello, F. 44, 94, 96, 281, 308
 Croft, Bruce 101, 276
 Crosson, B. 93, 284
 Cruz, N. F. 25, 298
 Cunningham, V. J. 52, 285, 301
 Cunningham, Vincent J. 28, 52, 269
 Cusack, Rhodri 43, 45, 272, 290
 Cutting, D. R. 99, 102, 276

 Dabringhaus, A. 44, 308
 Dacheille, Frank 85, 276
 Dale, Anders M. 27, 65, 276
 Damadian, R. 29, 276
 Dapretto, Mirella 44, 316
 Davatzikos, C. 94, 273
 Davatzikos, Christos 44, 48, 98, 276, 287
 David, A. S. 79, 274
 David, Antony 65, 273
 Davidon, W. C. 36, 277
 Davidowitz, Hanan 27, 307
 Davies, G. J. 86, 308
 Davis, G. 94, 280
 Davis, Gwendolyn 94, 293
 Davis, Karen D. 17, 277
 Davis, Matt H. 30, 46, 277

- Dawant, B. M. 43, 315
 Day, N. E. 61, 277
 de Jong, B. M. 98, 277
 de Jong, S. 72, 73, 138, 277
 de Jong, Sijmen 73, 277
 De Luna, Xavier 40, 277
 de Vos, K. 94, 96, 281
 Dearholt, D. W. 87, 310
 DeBellis, Robert 94, 283
 Decety, Jean 90, 98, 277
 Decombes, Xavier 41, 292
 Deerwester, Scott 99, 102, 277
 Deerwester, Scott C. 102, 277
 Dehaene, Stanislas 21, 22, 277, 312
 Deichmann, Ralf 30, 277
 Dekosky, Steven T. 123, 270
 Dellaert, Frank 97, 296
 Demb, J. B. 17, 277
 DeMers, D. 79, 138, 277
 Demonet, J. F. 33, 68, 272, 313
 Dempster, A. P. 136, 277
 Dence, C. 25, 280
 Denker, John S. 37, 70, 295
 Descombes, X. 48, 277
 Descombes, Xavier 51, 304
 Desmond, J. E. 17, 277
 D'Esposito, M. 25, 45, 48, 103, 108, 267
 D'Esposito, Mark 25, 46, 318
 Despres, D. 30, 314
 Desvignes, Michel 114, 309
 Devlin, Joseph T. 30, 46, 277
 Devlin, Joseph Thomas 30, 46, 277
 Dewey, S. L. 123, 310
 Deyoe, Edgar A. 68, 309
 Dhawan, A. P. 45, 318
 Diamantaras, K. I. 135, 293
 Diamantaras, Konstantinos I. 71, 73, 74, 79, 137, 277, 278
 Diehl, David J. 123, 270
 Diehr, P. 52, 297
 Dienel, G. A. 25, 298
 Diener, H. C. 122, 125, 290, 315
 Dietz, Alexander 85, 311
 Diligenti, Michelangelo 97, 278
 Ding, X. 52, 278
 Djeraba, Chabane 96, 278
 Dobelle, William H. 18, 278
 Dobkin, Jeffrey 123, 270
 Doddi, Srinivas 105, 278
 Dolan, R. J. 22, 48, 79, 279, 281, 282
 Dolan, Raymond 41, 280
 Doleisch, Helmut 91, 292
 Dom, Byron 96, 101, 274
 Domik, Gitta 88, 315
 Donaldson, David I. 25, 65, 292
 Dorosz, Jennifer L. 26, 308
 Doursat, R. 39, 135, 283
 Dowling, G. R. 99, 285
 Downar, Jonathan 17, 277
 Downham, D. Y. 40, 270
 Downs, H. 44, 294
 Downs, J. Hunter 95, 280
 Drury, Heather A. 45, 84, 89, 90, 278, 314
 DuBois, Richard M. 25, 275
 Duda, Andrzej 107, 315
 Duda, Richard O. 61, 278
 Duin, Robert P. W. 62, 63, 318
 Dumais, Susan T. 99, 102, 277
 Dumbreck, Andrew A. 88, 278
 Duncan, Gary H. 68, 69, 296
 Duncan, John 24, 278
 Duncan, K. 26, 289
 Dunn III, W. J. 73, 316
 Dunner, M. B. 17, 296
 Durso, F. T. 87, 310
 Dutschka, K. 125, 315
 Duyn, Jeff H. 46, 311
 Dziedzic, Mario 22, 45, 297
 Eastwood, James D. 28, 278
 Eaton III, Edward A. 104, 271
 Eckardt, C. 53, 278
 Eddy, W. F. 41, 279
 Eddy, William 45, 303
 Eddy, William F. 40, 93, 278, 283
 Edelman, R. R. 89, 315
 Eden, Guinevere F. 62, 98, 126, 314
 Eder, Derek 28, 268
 Einthoven, Willem 88, 278
 Ekatomdramis, D. 26, 290
 Elgar, Christian E. 28, 316
 Ellermann, J.M. 30, 303
 Emslie, Hazel 24, 278
 Engel, Stephen A. 26, 65, 272
 Engelen, Wolfgang 46, 304
 Eom, Sean B. 105, 278
 Erb, Michael 43, 52, 269, 278
 Erde, Steven M. 94, 283
 Erhard, P. 46, 288
 Erickson, Robert K. 28, 275, 314
 Eriksson, L. 44, 285, 310
 Eriksson, Lars 44, 289
 Ernst, T. 26, 279
 Ertl, Thomas 85, 311
 Evans, A. C. 84, 318
 Evans, Alan C. 41, 43–45, 48, 68, 69, 72, 74, 90, 93–95, 122, 123, 125, 269, 275, 276, 279, 282, 284, 290, 293–297, 299, 304, 317
 Everett, Peter 90, 283
 Everitt, B. S. 51, 279
 Everitt, Brian S. 62, 279
 Fabricius, Thomas 57, 279
 Fadili, M. J. 51, 279
 Fahlman, Scott E. 36, 279
 Faiss, J. H. 122, 290
 Falkman, Göran 87, 279
 Fantini, Sergio 31, 280
 Farah, Martha J. 90, 98, 279
 Farr, A. L. 109, 297
 Farris, Roy S. 105, 278
 Fawal, Houssam 114, 309
 Fayyad, Elias 85, 310
 Fazio, F. 109, 305
 Fell, J. 95, 303
 Fels, Gregor 88, 315
 Feng, Dagan 52, 279, 287
 Feng, Qingwen 87, 279
 Fiacco, A. V. 36, 279
 Filiaci, Mattia E. 31, 280
 Filzmoser, Peter 52, 279
 Findley, L. J. 123, 316
 Fischer, Michael J. 99, 315

- Fischman, Alan J. 22, 307
 Fisher, Ronald Aylmer 72, 153, 279
 Fisher, Scott S. 88, 299
 Fitzgerald, M. 41, 93, 278, 279
 Fitzpatrick, J. Michael 43, 315
 Flake, Gary 19, 155, 158, 295
 Flake, Gary W. 100, 292
 Flake, Gary William 105, 155, 158, 279, 295
 Flannery, Brian P. 36, 306
 Flaum, Michael 93, 284
 Fletcher, Jack M. 123, 310
 Fletcher, P. 22, 49, 55, 65, 95, 281, 282
 Fletcher, P. C. 79, 279
 Fletcher, Roger 36, 279
 Flitney, David 93, 311
 Foran, David 97, 276
 Foran, James 85, 274
 Ford, I. 48, 79, 299
 Ford, James 97, 279
 Forman, S. D. 41, 279
 Fortier, J. J. 74, 280
 Foster, D. 40, 280
 Fox, Peter T. 17, 18, 24, 25, 41, 44, 48, 49, 62, 81, 84, 94, 95, 97, 98, 123, 126, 136, 280, 292–294, 299, 304, 305, 317
 Frackowiak, Richard S. J. 22, 23, 25, 30, 40, 41, 43, 44, 47–49, 54, 55, 60, 65, 69, 71, 79, 80, 86, 95, 98, 99, 123, 125, 143, 277, 280–282, 288, 290, 298, 299, 306, 308, 316
 Frahm, J. 26, 292
 Frakes, W. B. 101, 280
 Fraley, Chris 50, 51, 280
 Franceschini, Maria Angela 31, 280
 Franckowiak, Richard S. J. 79, 279
 Frank, J. 30, 314
 Frank, Joseph A. 46, 311
 Frank, Lawrence R. 25, 26, 66, 71, 274
 Frazier, J. A. 143, 274
 Fredriksson, J. 44, 94, 96, 308
 Fredriksson, Jesper 94, 96, 281
 Freeman, Linton C. 87, 281
 Freeman, W. 55, 299
 Freeston, I. L. 31, 269
 Freitas, C. S. 24, 81, 95, 294
 Freitas, Catarina S. 94, 95, 294
 French, James C. 99, 104, 281
 Friderici, Angela D. 24, 298
 Frieder, Ophir 99, 100, 291
 Friedrich, R. 55, 281
 Frim, David M. 28, 275
 Friman, Ola 52, 79, 80, 281
 Friston, Karl J. 22, 24–26, 30, 38–49, 54, 55, 57, 60, 65, 67–69, 71, 72, 74, 79, 80, 90, 95, 99, 118, 123, 125, 136, 268, 269, 273, 279–282, 284, 288, 290, 291, 296, 299, 306, 312, 314, 316, 317
 Frith, Chris D. 22, 23, 41, 43, 44, 48, 49, 54, 55, 65, 71, 79, 80, 86, 95, 279–282, 298, 299, 308
 Fritsch, Christoph 18, 293
 Frost, J. A. 22, 26, 271
 Frostig, R. D. 30, 285
 Frutiger, Sally A. 39, 42, 51, 52, 78–80, 90, 91, 157, 269, 282, 291, 307
 Fu, Wenjiang J. 37, 282
 Fuchs, A. 55, 281, 282
 Fuchs, H. 86, 282
 Fuhita, H. 117, 295
 Fujimura, Kikuo 86, 282
 Fukuda, Hiroshi 123, 289
 Fukumizu, Michiro 31, 290
 Fulbright, Robert K. 123, 310
 Fullbright, Robert K. 78, 79, 299
 Fulton, R. 52, 279
 Fulton, Truxton K. 94, 296
 Furnas, George W. 99, 102, 277
 Fyfe, Colin 79, 293
 Gabrieli, J. D. 17, 277
 Gade, Anders 18, 21, 52, 91, 93, 158, 266, 282, 286
 Gadian, D. G. 23, 298
 Gale, William A. 137, 282
 Galsgaard, Klaus 88, 315
 Gansner, Emden R. 87, 283
 Gao, J.-H. 41, 317
 Gao, Jia-Hong 49, 317
 Garde, Ellen 45, 283
 Gardner, Daniel 94, 283
 Gardner, Esther P. 94, 283
 Garfield, Eugene 87, 99, 109, 115, 283, 311
 Garner, Line 93, 275
 Gaser, Christian 45, 283
 Gash, Don M. 55, 268
 Gaskil, M. 45, 318
 Gates, J. R. 17, 296
 Gati, Joseph S. 69, 300
 Gazzaniga, Michael S. 94, 96, 285
 Geiger, Bernhard 86, 125, 271, 283
 Geisser, S. 40, 283
 Geman, S. 39, 135, 283
 Genovese, Christopher R. 40, 45, 93, 278, 283, 303
 George, E. 40, 280
 George, John 31, 283
 George, M. S. 123, 283
 Georgopoulos, Apostolos P. 28, 307
 Gerber, Joanne 94, 296
 Gering, David T. 90, 283
 Gerring, Joan P. 98, 287
 Gerstein, George L. 27, 283
 Gessner, M. 95, 303
 Geyer, S. 44, 94, 96, 281, 308
 Geyer, Stefan 23, 283, 308
 Ghaffar, Omar 98, 296
 Ghahramani, Zoubin 64, 138, 307, 309
 Ghanayim, Nimr 18, 293
 Giannoylis, Irene 17, 277
 Gibson, David 96, 101, 104, 274, 284
 Gicquel, Sebastien 44, 306
 Gifford, David K. 107, 315
 Giles, C. Lee 19, 96, 97, 100, 104–106, 155, 158, 271, 278, 279, 284, 292, 295, 304
 Gill, Brad 25, 288
 Gill, D. S. 123, 283
 Girard, Franck 30, 298
 Gish, W. 109, 267
 Glantz, Stanton A. 120, 284
 Glaser, Daniel E. 40, 284
 Glass, T. G. 44, 294
 Glass, Thomas 95, 280
 Glenthøj, Birte 28, 268
 Glover, Eric 19, 100, 155, 158, 292, 295
 Glover, Eric J. 97, 284
 Glover, G. H. 17, 277
 Glover, Gary H. 25, 26, 65, 67, 272, 284
 Gokcay, D. 93, 284
 Gold, Sherri 93, 284

- Goldberg, I. E. 30, 293
 Goldfarb, D. 36, 284
 Goldwaithe, J. 18, 291
 Gollob, H. F. 53, 284
 Golub, Gene H. 136, 284
 Good, C. D. 23, 298
 Good, I. J. 137, 284
 Goodhill, Geoffrey J. 95, 284
 Gore, John C. 26, 78, 79, 123, 299, 308, 310
 Gori, Marco 97, 278
 Goualher, Georges Le 45, 284, 290
 Gouraud, H. 83, 284
 Goutte, Cyril 18, 25, 37, 39, 46, 48, 50–52, 57, 60, 65–68, 71, 93, 103, 157–187, 266, 284–286, 302, 303
 Govaert, G. 40, 51, 271
 Govaert, Gérard 40, 274
 Gower, J. C. 51, 54, 285
 Gowland, P. 26, 289
 Grabove, M. 93, 317
 Grady, C. L. 72, 79, 80, 288, 299
 Grafton, S. T. 43, 44, 93, 116, 317
 Grafton, Scott T. 123, 285
 Grant, Jonathan 120, 285
 Grasby, P. J. 40, 47, 282
 Grasby, P. M. 44, 300
 Grassi, F. 109, 305
 Grasso, Giorgio 96, 271
 Gratton, Enrico 31, 280
 Graufelds, C. J. 44, 308
 Graves, L. 25, 285
 Gray, J. A. 78–80, 273
 Green, A. A. 56, 313
 Green, Peter J. 60, 307
 Greenhouse, S. W. 40, 283
 Grefenstette, Gregory 100, 285
 Gregory, L. 78–80, 273
 Greitz, Torgny 44, 285, 289, 310
 Grethe, Jeffrey S. 94, 96, 285
 Greville, Thomas N. E. 136, 270
 Grèzes, Julie 90, 98, 277
 Griffith, B. C. 105, 316
 Grimson, W. Eric L. 90, 283
 Grinvald, Amiram 26, 30, 31, 285, 298, 314
 Gröller, Eduard 91, 292
 Grodd, Wolfgang 43, 52, 269, 278
 Groff, Jean-François 18, 270
 Grzesczuk, R. 28, 314
 Guelich, Scott 153, 285
 Guigon, E. 109, 305
 Gundavaram, Shishir 153, 285
 Gunderson, Brenda K. 78, 285
 Gunn, Roger N. 44, 52, 93, 269, 285, 300, 301
 Gunter, Thomas C. 24, 298
 Gusnard, Debra 22, 285
 Gustard, S. 46, 302
 Gyldensted, C. 30, 66, 303
- Ha, B. 93, 317
 Ha., Y. 79, 274
 Habib, Reza 98, 123, 296, 314
 Hadar, U. 125, 316
 Hagen, T. 28, 285
 Hagner, T. 95, 303
 Hahn, Erwin L. 109, 285
 Haken, H. 55, 281, 282
 Halber, M. 90, 287
- Hälikkälä, Sasu 87, 305
 Hall, P. A. V. 99, 285
 Halland, Debbie 43, 290
 Hammeke, T. A. 22, 26, 271
 Hannan, E. J. 39, 40, 286
 Hannerz, Jan 86, 288
 Hanquan, Y. 28, 288
 Hansen, Lars Kai 18, 25, 35, 37–40, 42, 44–46, 48–52, 54–57, 60–63, 65–68, 70, 71, 75, 78, 79, 84, 86, 89, 91, 93, 102, 103, 108, 109, 117, 120, 150, 157–266, 268, 269, 275, 279, 284–286, 288, 291, 292, 294, 295, 300–303, 305, 306, 312, 313
 Hansen, Per Christian 37, 137, 139, 286
 Hansen, Per Skafte 88, 286
 Hansen, Peter 93, 311
 Hansen, William W. 29, 271
 Hanson, Lars G. 26, 46, 271, 297
 Hanson, Stephen J. 96, 271
 Hansson, Per 87, 295
 Hanstock, C. 25, 306
 Hari, Riitta 44, 94, 96, 281, 308
 Harkness, B. 43, 315
 Harris, C. C. 43, 314
 Harshman, Richard 99, 102, 277
 Harshman, Richard A. 102, 277
 Hart, Peter E. 61, 278
 Harte, Robert 42, 299
 Hartigan, John A. 50, 286
 Hartvig, Niels Væver 46, 62, 287
 Hartvig, Niels Væver 41, 65, 81, 286
 Hasegawa, Osamu 51, 267
 Haselgrove, J. C. 25, 300
 Haslam, Jane 28, 52, 269
 Hassainia, Farid 84, 93, 298
 Hassibi, B. 37, 70, 287
 Hassibi, Babak 37, 287
 Hasson, Robert 28, 287
 Hastie, T. J. 69, 136, 287
 Hastie, Trevor 60, 287
 Hastie, Trevor J. 71, 317
 Hata, Noby 90, 283
 Hatano, Kentaro 123, 289
 Hawkes, D. J. 43, 315
 Haxby, J. V. 72, 79, 80, 129, 288, 298, 299
 Haxby, James V. 90, 96, 287, 289
 Hayamizu, Satoru 51, 267
 Haykin, Simon 135, 136, 287
 Hayton, Paul 62, 313
 Hazlett, Erin A. 45, 283
 Haznedar, M. M. 143, 287
 Healy, Matthew D. 93, 310
 Heather, J. D. 43, 44, 281
 Heeger, David J. 26, 65, 272
 Heimer, Lennart 23, 143, 287
 Heinze, H. J. 95, 303
 Hell, D. 28, 304
 Helmstaedter, Christoph 28, 316
 Hemler, P. F. 43, 315
 Hemmingsen, Ralf 28, 268
 Henn, S. 44, 310
 Hennig, J. 26, 279
 Henson, Richard Nevill Astley 40, 66, 68, 284, 287, 290
 Henzinger, Monika 105, 287
 Herholz, K. 90, 287
 Herskovits, Edward 44, 48, 276
 Herskovits, Edward H. 94, 98, 99, 287, 296, 314

- Hertz, John 36, 53, 73, 120, 135, 137, 287
Herzog, H. 27, 315
Herzog, Hans 24, 278
Hestenes, M. R. 36, 287
Hielscher, Andreas 31, 283
Hilgard, Ernest R. 125, 269
Hill, D. L. 43, 315
Hines, Michael S. 93, 310
Hinrichs, H. 95, 303
Hinterberger, Thilo 18, 293
Hinton, G. E. 37, 70, 287
Hinton, Geoffrey E. 79, 137, 270
Hintz-Madsen, Mads 35, 38, 63, 268, 294
Hirsch, Michael 93, 287
Hirsch, Traci B. 95, 280
Hlavac, Vaclav 81, 311
Ho, Cheng 31, 283
Ho, Dino 52, 287
Hochberg, Y. 41, 270
Hoerl, Arthur E. 35, 37, 65, 288
Hof, P. R. 143, 287
Hoffman, E. J. 29, 313
Hoffman, J. M. 43, 314
Hoffner, E. 125, 316
Hofman, Thomas 107, 275
Hofmann, Thomas 107, 288
Hoge, Richard D. 25, 288
Højen-Sørensen, Pedro A. d. F. R. 67, 288
Hold, Sherri 48, 268
Holder, David S. 28, 288
Hollander, E. 143, 287
Holm, Søren 25, 28, 117, 289, 295, 308
Holmberg, K.-G. 87, 305
Holmes, Andrew P. 25, 39, 40, 42, 46–49, 54, 57, 60, 65, 69, 71, 79, 80, 99, 117, 139, 140, 157, 200–221, 273, 281, 282, 288, 294, 299, 305, 306
Holmes, C. J. 45, 95, 275
Holmes, Colin J. 43, 93, 94, 116, 125, 276, 279, 317
Holte, S. 44, 285
Hopfinger, Joseph B. 42, 288
Hoppel, B. E. 30, 293
Hornak, Joseph P. 30, 288
Hornik, K. 74, 269
Horwitz, B. 48, 79, 288
Hotelling, Harold 52, 53, 71, 288
Houle, Sylvain 123, 314
Hounsfield, Godfrey N. 28, 288
Hovig, Eivind 87, 104, 290
Howard, David 123, 288
Howard, R. J. 79, 274
Howard, Robert 65, 273
Howseman, A. 25, 306
Howseman, A. M. 25, 60, 69, 299
Høy, Martin 77, 288
Hsieh, Jen-Chuen 86, 288
Hu, X. 17, 296
Hu, Xiaoping 26, 46, 288, 300, 317
Huang, Aileen 62, 280
Huang, Aileen Y. 62, 98, 280
Huang, C. L.-H. 46, 302
Huang, Y. 27, 315
Hubbard, Raymond 120, 268
Hübener, M. 30, 55–57, 312
Huber, J. P. 35, 289
Huber, W. 125, 315
Huggins, T. 123, 283
Hugh, Harlan M. 87, 289
Huiman, T. 26, 290
Hülsmann, Ernst 43, 278
Humphries, Jim 88, 299
Hund-Georgiadis, Margret 17, 289
Hunkin, N. M. 57, 312
Huole, S. 98, 314
Hurdal, Monica K. 90, 289
Hurvich, C. M. 40, 289
Hutchinson, M. 123, 310
Hutton, B. F. 43, 268
Hyde, J. S. 26, 271
Hyde, James S. 25, 46, 67, 68, 108, 269
Hykin, J. 26, 289
Iida, H. 43, 268
Imaizumi, Satoshi 123, 289
Inati, Souheil 94, 96, 285
Indefrey, Peter 22, 97, 98, 125, 289
Ingelman, L. 44, 308
Ingham, Janis C. 95, 280
Ingham, Roger J. 95, 280
Ingvar, David H. 24, 295
Ingvar, Martin 44, 86, 288, 289
Ingwersen, Peter 96, 104, 267, 289
Institute of Neurology, Wellcome Dept. of Cognitive Neurology 41, 289
Ionnides, Andreas A. 28, 289
Isbell Jr., Charles Lee 102, 289
Isensee, C. 125, 315
Ishai, Alumit 96, 289
Issa, B. 71, 292
Ito, U. 31, 298
Itoh, Kengo 123, 289
J. E. Adcock, Stephen M. Smith 17, 289
Jackson, Graeme 93, 267
Jackson, J. Edward 37, 53, 74, 289
Jaimes, Alejandro 89, 307
Jalinous, R. 31, 269
Janot, Nicolas 65, 273
Jarmasz, M. 48, 52, 269, 311
Jazczack, R. J. 43, 314
Jenike, M. A. 22, 143, 274, 307
Jenkins, I. H. 123, 316
Jenkinson, Mark 43, 93, 289, 311
Jensen, J. L. 46, 305
Jensen, Jens Ledet 62, 287
Jensen, Jesper James 85, 289
Jensen, Michael 28, 289
Jenssen, Tor-Kristian 87, 104, 290
Jerabek, Paul 95, 280
Jesmanowicz, A. 25, 46, 67, 68, 108, 269
Jezzard, P. 47, 65, 282
Jezzard, Peter 25, 290
Jobsis, F. F. 31, 290
Joenig, T. 28, 304
Joeri, P. 26, 290
Johnson, Brian 87, 290
Johnson, Debra L. 93, 284
Johnson, I. 26, 289
Johnson, J. D. 45, 318
Johnson, Keith A. 94, 290
Johnsrude, I. S. 23, 298
Johnsrude, Ingrid 43, 290
Jolesz, Ferenc 90, 283

- Joliot, Marc 44, 306
Jonathan, P. 75, 293
Jones-Gotman, M. 84, 318
Jones, Karen M. 62, 98, 126, 314
Jones, Richard A. 26, 290
Jones, T. 28, 52, 86, 290, 301, 308
Jones, Terry 28, 42, 52, 269, 299
Jöreskog, Karl Gustav 54, 63, 290
Joseffer, S. 123, 310
Josephs, Oliver 26, 46, 47, 65, 66, 71, 125, 143, 281, 282, 288, 290, 306, 308
Jueptner, M. 122, 290
Jung, T.-B. 57, 299
Jung, Tzyy-Ping 28, 57, 298, 299
Jungreis, C. A. 52, 297
- Kabani, N. J. 94, 276
Kabani, Noor 45, 290, 297
Kamei, A. 31, 290
Kamper, Lars 94, 95, 312
Kando, Noriko 104, 301
Kanno, I. 117, 295
Kanno, Iwao 43, 46, 47, 49, 55, 268, 291, 312
Kapur, S. 98, 314
Kapur, Shitu 123, 314
Karbe, H. 90, 287
Kardatzki, Bernd 52, 269
Karger, D. R. 99, 102, 276
Karhunen, K. 52, 290
Kashikura, K. 46, 47, 268
Kato, Hiroki 87, 99, 290
Kato, Takashi 123, 289
Kato, Toshinori 31, 290
Katz, Leonard 78, 79, 123, 299, 310
Kaufman, Arie 85, 276
Kaulisch, T. 95, 303
Kawashima, Ryuta 123, 289, 303
Kay, A. R. 30, 303
Kazman, Rick 87, 107, 274
Kedem, Z. M. 86, 282
Kelle, Olavi 84, 93, 290
Kelso, J. A. S. 55, 282
Kendall, D. G. 135, 291
Kennan, Richard P. 17, 291
Kennard, Robert W. 35, 37, 65, 288
Kennedy, D. N. 30, 293
Kennedy Jr, D. N. 30, 270
Kennedy, Phillip R. 18, 291
Kent, John T. 34, 35, 37, 49–51, 53, 54, 63, 64, 66, 70–75, 78, 86, 103, 104, 133–135, 137, 138, 145–148, 151, 153, 298
Kepler, Johannes 18, 291
Åkerman, S. 44, 308
Kershaw, Jeff 46, 47, 268, 291
Keskin, Can 87, 291
Kessler, J. 90, 287
Kessler, M. M. 99, 102, 291
Kessler, R. M. 43, 315
Kettenring, J. R. 78, 291
Ketter, T. A. 123, 283
Kety, S. S. 24, 291
Kevles, Bettyann Holtzmann 28, 120, 291
Kherif, F. 68, 313
Kidmose, Preben 57, 279
Kiebel, Stefan J. 42, 43, 291
Kiers, Henk A. L. 133, 138, 277, 291
Kikinis, Ron 90, 283
- Kim, Kwansik 85, 314
Kim, Seong-Gi 28, 30, 38, 48, 49, 157, 294, 303, 306, 307
Kinahan, Paul E. 153, 291
Kindermann, Sandra S. 57, 299
Kingston, Jeffrey H. 103, 291
Kiritani, Shigeru 123, 289
Kirkland, Kyle 27, 291
Kjell, Bradley 99, 100, 291
Kjems, Ulrik 39, 42, 44, 45, 61–63, 70, 120, 158, 241, 242, 286, 291, 300–302
Kleefield, J. 89, 315
Kleinberg, Jon 104, 284
Kleinberg, Jon M. 96, 99, 101, 104, 107, 274, 291
Kleinschmidt, A. 26, 292
Kling-Petersen, T. 84, 291
Klose, Uwe 43, 278
König, Andreas H. 91, 292
Knight, James Robert 100, 292
Knorr, U. 27, 315
Knowles, A. J. 71, 292
Knuth, Kevin H. 94, 283
Knutsson, Hans 52, 79, 80, 281
Kochi, K. 28, 304
Kochunov, Peter 44, 94, 95, 292, 294
Kochunov, Peter V. 94, 95, 294
Koelsch, Stefan 24, 298
Koenig, Michael E. D. 99, 105, 311
Kohl, M. 31, 302
Kohonen, Teuvo 54, 292
Koizumi, H. 31, 298
Kojima, Shozo 123, 289
Kolenda, Thomas 39, 49, 56, 57, 61, 63, 102, 117, 157, 158, 200–221, 286, 292, 294
Kollokian, V. 45, 94, 276, 292
Kollokian, Vasken 94, 275
Kolodny, Jonathan 24, 278
Komorowski, Jan 87, 104, 290
Konishi, Seiki 25, 65, 292
Koski, L. 98, 292
Koslow, Stephen 93, 287
Kosslyn, Stephen M. 123, 125, 292
Kostelec, Peter J. 94, 96, 285
Kostopoulos, G. 44, 94, 96, 281, 308
Kotchoubey, Boris 18, 293
Kotter, R. 95, 303
Kötter, Rolf 94, 95, 312
Koukko, M. 28, 304
Koulis, T. 41, 317
Koutsofios, Eleftherios 87, 292
Krabbe, Katja 45, 283
Kreeger, Kevin 85, 276
Kristensen, Benny Lønstrup 71, 297
Kristeva, Romyana 27, 315
Krogh, Anders 36, 53, 73, 120, 135, 137, 287
Krovetz, Robert 19, 155, 158, 295
Kruger, Andries 19, 100, 155, 158, 292, 295
Krüger, G. 26, 292
Krugger, F. 48, 277
Krugger, Frithjof 41, 51, 65, 69, 292, 304, 307
Krugger, Frithof 69, 292
Krutchkoff, R. G. 33, 292
Krzanowski, W. J. 75, 79, 293
Kubik, Stefan 23, 303
Kübler, Andrea 18, 293
Kukla, James M. 100, 311
Kulick, Todd 85, 293

- Kullback, S. 139, 293
Kumar, Ravi 104, 273
Kumar, S. Ravi 96, 101, 274
Kung, S. Y. 135, 293
Kung, Sun-Yuan 71, 73, 74, 79, 137, 277, 278
Kuo, Eddy 86, 282
Kuroda, Kagayaki 30, 293
Kustra, Rafal 75, 145, 293
Kuwabara, T. 25, 306
Kwan, Chun L. 17, 277
Kwan, Remi K.-S. 94, 275, 293
Kwong, K. K. 30, 293
- Lacadie, Cheryl 78, 79, 299
Læreid, Astrid 87, 104, 290
Lai, Pei Link 79, 293
Laird, N. M. 136, 277
Lakshminarayan, Kamakshi 90, 91, 94, 307
Lammertsmaa, A. A. 48, 281
Lamping, John 87, 293
Lamping, John O. 87, 293
Lancaster, Jack L. 24, 41, 44, 45, 49, 62, 81, 94, 95, 97, 98, 126, 136, 272, 280, 292–294, 299, 317
Landauer, Thomas K 102, 277
Lange, Nicholas 18, 47–49, 52, 54, 65, 69–71, 78, 79, 93, 117, 157, 158, 200–221, 266, 286, 294, 300, 301, 312, 313
Lange, Nichols 78, 312
Lange, Nick 49, 157, 294
Lankipalli, B. R. 44, 294
Laplane, D. 109, 294
Larkey, Leah S. 140, 294
Larsen, Jan 35, 38–40, 54, 56, 57, 61, 63, 69, 70, 150, 157, 158, 189–199, 268, 284, 286, 294, 313
Larson, Ray R. 105, 295
Larsson, Henrik B. 25, 308
Larsson, Henrik B. W. 17, 26, 30, 45, 51, 52, 272, 283, 316
Larsson, Stig 87, 295
Lassen, Niels A. 24, 117, 295
Lauffer, R. B. 30, 315
Lauterbur, Paul G. 29, 295
Lautrup, Benny 39, 55, 70, 71, 295, 300, 301
Law, Ian 22, 25, 26, 28, 42, 48, 51, 52, 55, 70, 71, 84, 98, 117, 127, 128, 140, 157, 269, 271, 295, 300, 301, 308
Lawrence, Steve 19, 96, 97, 100, 104–106, 155, 158, 271, 278, 279, 284, 292, 295, 304
Le Bihan, D. 30, 314
Le Cun, Y. 36, 270
Le Cun, Yann 37, 70, 295
Le, Tuong H. 46, 288
Le, Tuong Huu 26, 317
Leach, M. O. 30, 295
Leahy, R. M. 28, 42, 269, 301, 309
Leahy, Richard 42, 309
Leblanc, R. 122, 123, 295
Ledberg, A. 44, 308
Ledberg, Anders 41, 52, 79, 267, 296
Lee, Daniel D. 49, 55, 296
Lee, T. M. 30, 303
Lee, Te-Won 56, 57, 296, 299
Leehan, Joaquín Azpiroz 84, 93, 298
Leen, Todd K. 37, 296
Lefkowitz, D. 52, 297
Lehmann, D. 28, 304
Lehnertz, Klaus 28, 316
Leibler, R. A. 139, 293
Leibovici, Didier 93, 311
Leibovici, Didier G. 78, 296
Lemoine, D. 43, 315
Leonard, C. M. 93, 284
Lepage, Martin 98, 296
Leschinger, A. 95, 303
Leshner, S. 31, 285
Leth, H. 26, 30, 272
Letovsky, Stan I. 94, 296
Levelt, Willem J. M. 22, 97, 98, 125, 289
Levenberg, K. 36, 296
Levesque, M. 129, 310
Levin, Asriel U. 37, 296
Levitt, J. B. 57, 309
Lewis, Bobbi K. 46, 311
Lewis, P. S. 28, 301
Lex, Ulrike 17, 289
Li, Henry H. 44, 48, 276
Liao, C. 93, 317
Liao, Chuanhong H. 68, 69, 296
Liao, Y.-K.C. 97, 296
Liddle, P. F. 41, 48, 49, 55, 95, 281, 282
Lieke, E. 30, 285
Lin, J. C. 17, 296
Lindley, D. V. 37, 296
Liotti, Mario 94, 95, 294
Liow, J. S. 42, 43, 55, 269, 312
Lipman, D. J. 109, 267
Lipschutz, Brigitte 46, 296
Lipshutz, Brigitte 46, 296
Liptrot, Matthew George 18, 39, 46, 51, 52, 57, 93, 158, 266, 284, 286, 313
Liseo, Brunero 139, 270
Liu, C. 89, 315
Liu, C. S. 25, 300
Liu, G. T. 25, 300
Liu, Yanxi 97, 296
Lloyd, Dan 98, 297
Lobaugh, N. J. 79, 297
Lochbaum, Karen E. 102, 277
Loenneker, Th. 26, 290
Loève, M. 52, 297
Loh, Wei-Liem 78, 297
Lohmann, Gabriele 93, 297
Longstaff, Alan 24, 297
Longstreth Jr, W. T. 52, 297
Lorensen, Bill 85, 310
Lorensen, William E. 85, 275, 297
Lou, H. C. 26, 271
Lou, Hans C. 17, 26, 30, 271, 272
Lowe, Mark J. 22, 45, 48, 297
Lowry, Oliver H. 109, 297
Ludke, Siegwalt 85, 275
Luhn, H. P. 102, 297
Luknowsky, David C. 69, 300
Lund, J. S. 57, 309, 310
Lund, Torben E. 26, 46, 271, 297
Lundberg, Peter 52, 79, 80, 281
Lundsager, Bent 71, 297
Lunin, Lois F. 105, 135, 297
Luotonen, A. 18, 270
Lurito, Joseph T. 22, 45, 297
Luu, Phan 98, 274
Luukkonen, Terttu 87, 305
- MacDonald, B. 129, 310
MacDonald, David 41, 45, 290, 297, 317

- MacFall, J. R. 43, 314
 MacGregor, John F. 66, 71, 74, 76, 149, 273
 Maciunas, R. J. 43, 315
 MacJones, Paul 96, 271
 MacKay, David J. C. 69, 70, 135, 136, 297, 298
 Mackinlay, Jock D. 87, 308
 Macko, K. A. 95, 300
 MacQueen, J. 50, 298
 Macready, W. G. 37, 316
 Madoff, David C. 123, 270
 Madsen, Henrik 47, 298
 Madsen, Peter Lund 25, 26, 271, 298
 Maes, F. 43, 315
 Maes, Frederik 43, 298
 Maess, Burkhard 24, 298
 Maghoul, Farzin 104, 273
 Maguire, Eleanor A. 23, 298
 Mailis, Angela 17, 277
 Maintz, J. B. 43, 315
 Maintz, J. B. Antoine 41, 298
 Maisog, Jose M. 65, 307
 Maisog, José Ma. 48, 298
 Makedon, Fillia 97, 279
 Makeig, Scott 28, 57, 298, 299
 Maki, A. 31, 298
 Malak, J. 31, 302
 Malandain, G. 43, 315
 Malikovic, A. 23, 268
 Malikovic, Aleksander 23, 308
 Malonek, D. 26, 298
 Manelfe, C. 33, 272
 Mangin, Jean-François 93, 275
 Manolakis, Dimitris G. 33, 306
 Manolio, T. A. 52, 297
 Mansfield, P. 26, 289
 Manton, D. J. 71, 292
 Marangell, L. 123, 283
 Marathe, Madhav V. 105, 278
 Marcar, Valentine L. 30, 298
 Marchal, G. 43, 298
 Marchini, Jonathan L. 46, 298
 Marchione, Karen E. 78, 79, 299
 Mardia, Kantilal Varichand 34, 35, 37, 49–51, 53, 54, 63, 64, 66, 70–75, 78, 86, 103, 104, 133–135, 137, 138, 145–148, 151, 153, 298
 Markowitsch, Hans J. 123, 314
 Marquardt, D. W. 36, 298
 Marrelec, Guillaume 66, 118, 298
 Marrett, S. 41, 48, 317
 Marrett, Sean 25, 288
 Marrin, Chris 84, 274
 Marriott, F. H. C. 51, 78, 298
 Martens, Harald 71–73, 76, 77, 288, 298
 Martidis, Adam 123, 270
 Martin, A. 129, 298
 Martin, Alex 96, 289
 Martin, Charles 95, 280
 Martin, E. 26, 290
 Martin, Ernst 30, 298
 Martin, Ken 85, 310
 Martin, Richard F. 23, 87, 94, 143, 272
 Martinelli, G. 36, 307
 Martínez, Alfonso Martí 84, 93, 298
 Martinez, Dominique 63, 299
 Masaryk, T. 52, 278
 Massy, W. F. 37, 66, 138, 299
 Mata, Daniel Martí. 84, 93, 298
 Matsumura, Akira 30, 293
 Matthews, Julian 42, 299
 Matthews, P. M. 17, 289
 Matthews, Paul M. 30, 46, 277
 Matthews, Vincent P. 22, 45, 297
 Mattle, H. P. 89, 315
 Maurer Jr, C. R. 43, 315
 Mayanagi, Y. 31, 298
 Mayberg, H. 44, 294
 MayerKress, G. 55, 299
 Mayhew, J. E. W. 57, 309, 310
 Mazoyer, B. 44, 51, 94, 96, 279, 281, 308
 Mazoyer, B. M. 41, 48, 306
 Mazoyer, Bernard 44, 306
 Mazoyer, Nathalie 44, 306
 Mazziotto, J. C. 42, 269, 309
 Mazziotto, John C. 30, 41, 43, 44, 93, 95, 116, 123, 280, 282, 285, 294, 299, 316, 317
 McAvoy, M. P. 40, 303
 McCain, Katherine W. 99, 104, 105, 299, 301
 McCain, Kathrine W. 105, 106, 135, 299, 316
 McCarthy, Gregory 27, 28, 98, 267
 McCarthy, W. V. 75, 293
 McColl, J. H. 48, 299
 McCormick, G. P. 36, 279
 McCrory, S. J. 79, 299
 McDowall, Ian E. 88, 299
 McGonigle, David J. 25, 60, 69, 299
 McGuire, P. K. 22, 299
 McIntosh, Anthony Randal 72, 79, 80, 297, 299
 McIntyre, M. 46, 52, 309
 McKeown, Martin J. 57, 299
 McKinsty, R. C. 30, 270
 McLachlan, G. J. 51, 299
 McLaughlin, Gerald W. 105, 269
 McLoughlin, N. 57, 309, 310
 Mead, R. 36, 302
 Meade, Jeffrey 90, 301, 318
 Meador, K. J. 31, 307
 Mechelli, A. 65, 71, 282
 Medina, Juan Manuel Reyes 84, 93, 298
 Medl, Attila 97, 276
 Meer, Peter 97, 276
 Megalooikonomou, Vasileios 98, 287
 Megalooikonomou, Vasilis 97, 279
 Meininger, V. 109, 294
 Mellers, John 65, 273
 Mencl, W. Einar 78, 79, 299
 Mendelzon, Alberto 107, 300
 Mendelzon, Alberto O. 107, 307
 Menon, R. 30, 303
 Menon, R. S. 26, 300
 Menon, Ravi S. 69, 300
 Mentzel, Heiko 93, 297
 Merkle, H. 30, 303
 Merlo, M. C. 28, 304
 Metropolis, N. 137, 300
 Meyer, E. 84, 122, 123, 295, 318
 Meyer, J. H. 44, 300
 Meyer-Lindenberg, Andreas 45, 300
 Michalos, Antonios 31, 280
 Miezín, F. M. 81, 280
 Miezín, Francis M. 123, 273, 276
 Miki, A. 25, 300
 Mikiten, S. A. 94, 280

- Mikiten, Shawn 94, 293
 Mikiten, Shawn A. 94, 95, 293, 294
 Mikulis, David J. 17, 277
 Miller, David J. 96, 300
 Miller, Gerald A. 94, 296
 Miller, H. D. 103, 105, 276
 Miller, Perry L. 93, 310
 Milner, Brenda 123, 304
 Milton, John 28, 275, 314
 Mintun, M. 84, 95, 123, 305
 Mintun, M. A. 25, 41, 48, 279, 280
 Mintun, Mark A. 123, 270
 Miranda, Maria J. 17, 26, 272
 Mirsky, Jason S. 93, 310
 Mishkin, M. 79, 95, 288, 300
 Mitra, Sanjit Kumar 65, 67–69, 73, 300
 Miulli, D. E. 17, 296
 Miura, S. 117, 295
 Mockus, A. 93, 278
 Moeller, J. R. 48, 55, 300
 Mohlberg, H. 23, 268
 Mohlberg, Hartmut 23, 24, 268
 Mohr, C. M. 93, 284
 Molgedey, Lutz 56, 300
 Møller, Martin Fodsslette 36, 300
 Moody, John E. 37, 40, 296, 300
 Moonen, C. T. W. 30, 314
 Moonier, P. A. 42, 299
 Moore, C. 22, 282
 Moore, Melody M. 18, 291
 Moore, R. 26, 289
 Mørch, Niels Jacob Sand 39, 52, 54, 55, 65, 68, 70, 71, 157, 159, 286, 295, 300, 301, 303, 313
 Moreno, Jacob Levi 87, 301
 Mori, Koichi 123, 289
 Morosan, P. 23, 301
 Morosan, Patricia 43, 290
 Morrell, M. J. 17, 277
 Morris, R. G. 78–80, 273
 Morris, Theodore A. 99, 104, 105, 301
 Mortensen, E. L. 45, 283
 Moscovitch, M. 98, 314
 Moser, Ewald 52, 279
 Mosher, J. C. 28, 301
 Moss, Helen E. 30, 46, 277
 Mueller, D. Timmann S. P. 122, 290
 Mueller, Klaus 90, 301, 318
 Muirhead, Robb J. 78, 285
 Muley, Suraj 39, 42, 291
 Mullani, N. A. 29, 313
 Müller, Karsten 93, 297
 Muller, S. 125, 315
 Mummery, C. J. 71, 273
 Munk, O. L. 43, 301
 Munzner, Tamara 87, 301
 Murata, Nobora 40, 301
 Murata, Noboru 40, 301
 Murtagh, B. A. 36, 301
 Myers, E. W. 109, 267
 Myers, R. 44, 52, 300, 301

 Nabavi, Arya 90, 283
 Nadkarni, Parkash M. 93, 310
 Næs, Tormod 71–73, 76, 298
 Najork, Marc A. 83, 301
 Nakai, Toshiharu 30, 293
 Nakamura, Akinori 123, 289
 Nakamura, Katsuki 123, 289
 Nakayama, Takehiro 87, 99, 290
 Nanba, Hidetsugu 104, 301
 Napel, S. 43, 315
 Narial, T. 52, 314
 Narin, Francis 99, 105, 274
 Neal, Radford M. 35, 70, 302
 Nebeling, B. 27, 315
 Neelin, P. 41, 43–45, 48, 269, 276, 317
 Nelder, J. A. 36, 302
 Nelson, C. A. 17, 296
 Nemprempre, Chanathip 107, 315
 Nenadic, Igor 45, 283
 Netsiri, C. 46, 302
 Neufang, M. 31, 302
 Nevill-Manning, Craig C. 96, 302
 Newell, Fiona N. 24, 278
 Nguyen, S. 94, 293
 Nicholas, Charles 100, 311
 Nichols, Thomas E. 48, 57, 139, 140, 305
 Nicholson, Keith 17, 277
 Nickerson, Dan 94, 95, 294
 Nickerson, Daniel 95, 294
 Nickles, R. J. 28, 289
 Nielsen, Finn Årup 18, 19, 25, 35, 37, 38, 48–52, 54, 55, 57, 60–63, 65, 66, 68, 71, 75, 84–86, 89, 91, 93, 103, 108, 109, 117, 120, 155, 157–266, 269, 275, 284–286, 294, 295, 302–304, 306, 313
 Nielsen, H. 18, 270
 Nielsen, Hans Bruun 37, 303
 Nielsen, Jakob 91, 303
 Nikerson, Dan 95, 294
 Noble, J. A. 46, 270
 Nobre, Anna 27, 28, 267
 Noll, D. C. 41, 279
 Noll, Douglas C. 26, 40, 45, 93, 153, 278, 283, 291, 303, 314
 Nooruddin, F. S. 86, 303
 Nordell, B. 44, 310
 North, Stephen C. 87, 283, 292, 303
 Northoff, G. 95, 303
 Northway, M. L. 87, 303
 Noth, J. 125, 315
 Novotny, E. 25, 306
 Nowinski, Wieslaw L. 90, 303
 Nyberg, L. 79, 80, 299
 Nyberg, Lars 98, 274, 296

 Obermayer, K. 57, 310
 Obermayer, Klaus 30, 55–57, 309, 312
 Obrig, H. 31, 302
 O'Craven, K. 54, 79, 312
 Ogawa, S. 26, 300
 Ogawa, Segi 30, 303
 Ogilvie, Paul 140, 294
 Ohlsson, Mattias 70, 303
 Ohyu, Junkou 31, 290
 Okada, Tomohisa 30, 293
 Okumura, Manabu 104, 301
 O'Leary, Daniel 48, 268
 Ollinger, J. M. 40, 65, 67, 303
 Omlin, Christian 100, 292
 Ono, Michi 23, 303
 O'Reilly, G. V. 89, 315
 Oshio, Koichi 30, 293
 Østergaard, L. 30, 66, 303

- O'Sullivan, Brendan T. 123, 303
 Otto, Thomas 30, 55–57, 312
 Owen, Adrian M. 123, 304
 Ozaki, T. 31, 290
- Pack, A. 25, 285
 Packard, Martin 29, 271
 Page, Lawrence 105, 272, 304
 Paik, Chul H. 94, 296
 Palm, Güther 52, 269
 Palmer, Richard G. 36, 53, 73, 120, 135, 137, 287
 Palomero-Gallagher, Nicola 23, 318
 Palubinskas, Gintautas 51, 304
 Pan, Hong 46, 304
 Pardo, José V. 17, 123, 304
 Parekh, B. A. 123, 283
 Parisi, Anthoy 84, 270
 Park, Trevor 46, 304
 Parrish, T. 46, 288
 Parsons, Lawrence M. 62, 97, 98, 126, 136, 280
 Parsons, Mally 95, 294
 Parzen, Emanuel 60, 138, 304
 Pascher, R. 84, 291
 Pascual-Marqui, Roberto D. 28, 304
 Passingham, R. E. 143, 308
 Pastoor, S. 88, 311
 Pastoor, Siegmund 88, 304
 Patterson, Karalyn 123, 288
 Paulesu, E. 22, 86, 299, 308
 Pauling, L. 30, 304
 Paulson, Olaf B. 18, 25, 26, 39, 42, 51, 52, 54, 70, 71, 84, 93, 117, 127, 128, 157, 158, 189–199, 266, 269, 271, 286, 295, 300, 301, 308, 313
 Paus, Tomáš 98, 304
 Paus, Tomáš 98, 292
 Pazaglia, P. J. 123, 283
 Pearce, S. C. 135, 304
 Pearson, Helen 104, 304
 Pearson, K. 52, 53, 138, 304
 Pedersen, J. 44, 308
 Pedersen, Morten With 49, 71, 294, 304
 Peitersen, Birgit 17, 26, 30, 272
 Pekar, J. 129, 315
 Pelizzari, C. A. 43, 304, 314
 Peltier, Scott J. 30, 304
 Penderson, J. O. 99, 102, 276
 Pennock, David M. 19, 155, 158, 295, 304
 Penny, William D. 40, 284
 Penrose, Roger 137, 304
 Perani, F. Lacquaniti D. 109, 305
 Perelmouter, Juri 18, 293
 Perkel, D. H. 27, 283
 Persson, Olle 87, 305
 Pesce, Mark 84, 270
 Peters, Terry M. 45, 95, 275, 276
 Petersen, Kim Spetzler 56, 57, 305
 Petersen, Niels Væver 46, 305
 Petersen, S. E. 98, 273
 Petersen, Steve E. 123, 273
 Petersen, Steven E. 84, 95, 123, 276, 305
 Peterson, Carsten 70, 104, 271, 303, 305
 Petersson, Karl Magnus 48, 57, 139, 140, 305
 Petit, Laurent 90, 287
 Petre, V. 93, 317
 Petrides, Michael 123, 304
 Petroff, O. 25, 306
- Pettigrew, Karen 41, 314
 Phatak, Aloke 73, 277
 Phelps, Michael E. 29, 305, 313
 Philips, D. L. 37, 305
 Philips, Jacquie 71, 282
 Philips, Micheal D. 22, 45, 297
 Philipsen, Peter Alshede 42, 305
 Phong, Bui Tuong 83, 305
 Pi, Hong 70, 303, 305
 Picard, N. 98, 112, 305
 Pieper, Steven D. 88, 299
 Piepgras, U. 28, 285
 Pietrzyk, U. 90, 287
 Pike, G. Bruce 25, 94, 288, 293
 Pinel, P. 21, 277
 Pirolli, Peter 87, 293
 Pirolli, Peter L. 104, 305
 Pitkow, James E. 104, 305
 Pitkow, James Edward 18, 105, 305
 Pizzi, N. 52, 311
 Pizzi, Nicolino J. 93, 305
 Pohl, G. 90, 270
 Polak, E. 36, 305
 Poldrack, Russ 81, 306
 Poline, Jean-Baptiste 40–44, 47–49, 54, 57, 68, 69, 71, 72, 74, 79–81, 93, 99, 139, 140, 273, 275, 281, 282, 291, 296, 305, 306, 317
 Poljak, B. T. 36, 306
 Pollermann, Bernd 18, 270
 Pomerleau, Dean A. 63, 70, 269, 306
 Poncelet, B. P. 30, 293
 Poppelwell, D. 44, 94, 96, 308
 Popplewell, D. 94, 96, 281
 Porrill, J. 57, 312
 Porter, M. F. 99, 101, 306
 Porter, N. R. 57, 312
 Posner, M. I. 84, 95, 123, 305
 Posner, Michael I. 17, 81, 98, 274, 306
 Posse, S. 44, 310
 Post, R. 123, 283
 Pound, R. V. 29, 306
 Powell, M. J. D. 36, 279
 Press, William H. 36, 306
 Price, Cathy J. 22, 24, 30, 46, 65, 71, 125, 277, 282, 296, 306, 314
 Price, M. Andrew 140, 294
 Price, Pat 42, 299
 Prichard, J. 25, 306
 Proakis, John G. 33, 306
 Procyk, E. 45, 284
 Puce, Aina 27, 28, 98, 267
 Pugh, Kenneth R. 78, 79, 123, 299, 310
 Purcell, Edward Mills 29, 306
 Purdon, Patrick L. 46, 47, 306
 Purushotham, Archana 38, 48, 157, 306
- Quinn, B. G. 39, 40, 286
 Quinton, Olivier 44, 306
- Rabe-Hesketh, S. 78–80, 273
 Rabe-Hesketh, Sophia 65, 273
 Rabi, Isidor Isaac 29, 307
 Rademacher, J. 23, 301
 Rafiei, Davood 107, 300, 307
 Raftery, Adrian E. 50, 51, 280
 Ragazzani, S. 36, 307

- Raghavan, Prabhakar 96, 101, 104, 273, 274, 284
 Raghavan, Ragnu 90, 303
 Raichle, Marcus E. 17, 22, 25, 48, 81, 84, 95, 123, 273, 280, 285, 304–306
 Rainey, L. 62, 280
 Rainey, Lacey 62, 98, 280
 Rainey, Lacy 94, 95, 294
 Rainey, Lacy H. 95, 294
 Rainey, Lance 24, 81, 95, 294
 Rajagopalan, Sridhar 96, 101, 104, 273, 274
 Rajapakse, Jagath C. 65, 307
 Randall, R. J. 109, 297
 Ranjeva, J. P. 68, 313
 Rao, A. 28, 288
 Rao, A. Ravishankar 89, 307
 Rao, C. R. 74, 307
 Rao, Ramaba B. 87, 293
 Rao, Ramana 87, 293
 Rao, S. 17, 317
 Rao, S. M. 22, 26, 271
 Rapoport, S. I. 79, 288
 Rasmussen, Carl Edward 37, 38, 60, 67, 69, 138, 286, 288, 307, 316
 Rasmussen, T. 17, 315
 Rauch, S. L. 143, 274
 Rauch, Scott L. 22, 123, 292, 307
 Rauschecker, J. P. 129, 315
 Ravi, S. S. 105, 278
 Ray, P. G. 31, 307
 Raz, J. 25, 300
 Rebbechi, Donovan 96, 271
 Reber, Arthur S. 125, 307
 Rector, David 31, 283
 Reed, Todd 96, 302
 Rees, Geraint 26, 71, 273, 282
 Reeves, C. M. 36, 279
 Rehm, Kelly 70, 78, 90, 91, 94, 289, 301, 307, 312
 Resnick, Susan M. 44, 48, 276
 Revenu, Marinette 114, 309
 Rezza, M. 54, 79, 312
 Ribiere, G. 36, 305
 Richardson, Sylvia 60, 307
 Richter, A. 95, 303
 Richter, Wolfgang 28, 48, 307, 311
 Ricotti, L. P. 36, 307
 Rijntjes, M. 122, 125, 290, 315
 Rinberg, Dmitry 27, 307
 Ripley, Brian D. 34, 36–38, 40, 46, 60, 61, 75, 93, 134, 298, 307, 311
 Rissanen, J. 40, 307
 Ritter, F. J. 17, 296
 Robb, Richard A. 83, 307
 Roberts, Gareth O. 105, 272
 Roberts, Stephen 62, 63, 308
 Roberts, Stephen J. 62, 63, 308
 Robertson, George G. 87, 308
 Robertson, Stephen E. 102, 308
 Robson, Matthew D. 26, 308
 Rockmore, Daniel 94, 96, 285
 Rögvaldsson, Thorsteinn 70, 303
 Roland, Per E. 23, 41, 44, 94, 96, 123, 281, 296, 303, 308, 310
 Roman, D. D. 17, 296
 Ropella, Kristina M. 68, 309
 Rorden, Chris 23, 90, 308
 Rosebrough, N. J. 109, 297
 Rosen, Bruce R. 30, 52, 54, 65, 66, 68, 79, 143, 157, 159, 270, 274, 286, 293, 303, 312, 313, 315
 Rosen, S. D. 86, 308
 Rosenbaum, Howard 96, 311
 Rosenblatt, F. 138, 308
 Rosenblatt, Murray 60, 308
 Rosenbluth, A. W. 137, 300
 Rosenbluth, M. N. 137, 300
 Rosenqvist, G. 44, 310
 Rosenthal, Jeffrey S. 105, 272
 Rosenthal, R. 97, 308
 Rossetti, Y. 109, 305
 Rostrup, Egill 17, 25, 26, 30, 39, 42, 45, 46, 48, 50–52, 54, 57, 60, 65, 67, 68, 84, 103, 117, 127, 128, 157, 159, 175–187, 189–199, 271, 272, 283–286, 288, 295, 303, 308, 313
 Rothman, D. 25, 306
 Rottenberg, David A. 18, 39, 42, 43, 48, 49, 51, 52, 54, 55, 71, 78–80, 90, 91, 93, 94, 157, 158, 266, 269, 282, 286, 289, 291, 300, 301, 307, 309, 312
 Rouquette, S. 47, 282
 Rowe, James B. 143, 308
 Roweis, Sam 54, 64, 74, 139, 308, 309
 Roy, C. S. 24, 309
 Royackkers, Nicolas 114, 309
 Ruan, S. 51, 279
 Rubin, Donald B. 54, 136, 277, 309
 Rugg, Michael D. 40, 65, 68, 281, 284, 287
 Ruggieri, P. 52, 278
 Ruhe, A. 73, 316
 Ruppert, David 46, 304
 Rus, Daniela 94, 96, 285
 Rusinek, H. 123, 310
 Russell, Richard P. 30, 46, 277
 Ruttiman, Urs R. 46, 311
 Rydmark, M. 84, 291
 Ryner, Lawrence 52, 269
 Saad, Ziad 68, 270
 Saad, Ziad S. 68, 309
 Saarimaki, A. 123, 310
 Sahami, Mehran 99–102, 309
 Sakamoto, Toshio 30, 293
 Salton, Gerard 99–102, 309
 Salzberg, B. M. 31, 285
 Sampson, Geoffrey 137, 282
 Sanders, G. 30, 314
 Sandor, Stephanie 42, 309
 Sargent, R. W. H. 36, 301
 Sarukkai, Ramesh R. 99, 309
 Savage, Cary R. 22, 307
 Savoy, Robert 39, 49, 52, 54, 65, 68, 71, 79, 117, 157, 159, 189–221, 286, 294, 301, 303, 312, 313
 Savoy, Robert L. 49, 71, 294
 Saykin, Andrew J. 97, 279
 Scarth, Gordon B. 46, 52, 309
 Schaltenbrand, Georges 90, 309
 Schaper, Kirt A. 42, 49, 55, 78, 90, 91, 94, 269, 289, 307, 309, 312
 Schapiro, M. B. 79, 288
 Scharf, L. L. 74, 309
 Schauer, Margarete 18, 293
 Scheich, H. 95, 303
 Scheiderman, Ben 105, 272
 Schiessl, I. 30, 55–57, 309, 310, 312
 Schlaug, G. 27, 315
 Schleicher, Axel 23, 24, 44, 94, 96, 268, 281, 283, 301, 308

- Schlösser, R. 123, 310
 Schmidt, C. F. 24, 291
 Schmidt, Kathleen 41, 314
 Scholz, U. 31, 302
 Schormann, Thorsten 23, 44, 94, 96, 268, 281, 301, 308, 310
 Schreiner, C. E. 56, 269
 Schuster, H. G. 56, 300
 Schvaneveldt, Roger W. 87, 310
 Schwartz, Candy 96, 310
 Schwarz, G. 40, 310
 Schlagenhauf, F. 95, 303
 Schroeder, William 85, 310
 Secret, A. 18, 270
 Seidman, L. J. 143, 274
 Seitz, Rüdiger J. 24, 27, 44, 278, 310, 315
 Sejnowski, Terrence J. 28, 57, 270, 298, 299
 Selden, Catherine Ross 97, 310
 Sen Gupta, Ashis 78, 310
 Senda, M. 52, 314
 Seppa, M. 44, 94, 96, 281, 308
 Sereno, Martin I. 27, 276
 Sergeant, J. 129, 310
 Sethi, Ravi 100, 267
 Seung, H. Sebastian 49, 55, 296
 Shahrababaki, Khatoun 88, 315
 Shallice, T. 79, 279
 Sham, Pak 65, 273
 Shankweiler, Donald P. 123, 310
 Shanno, D. F. 36, 310
 Shannon, C. E. 138, 310
 Shao, J. 39, 310
 Shattuck, D. W. 42, 269, 309
 Shaywitz, Bennett A. 78, 79, 123, 299, 310
 Shaywitz, Sally E. 78, 79, 123, 299, 310
 Shear, P. K. 17, 277
 Shekhar, Raj 85, 310
 Sheldon, Mark A. 107, 315
 Shen, Li 97, 279
 Shepherd, Gordon M. 24, 27, 93, 310
 Sherrington, C. S. 24, 309
 Shibata, R. 39, 310
 Shipp, S. 98, 277
 Shneiderman, Ben 87, 290, 310
 Shulman, Gordon 22, 285
 Shulman, Gordon L. 65, 67, 123, 276, 303
 Shulman, R. 25, 306
 Sicotte, Nancy L. 44, 116, 316, 317
 Sidtis, John J. 48, 49, 51, 52, 54, 55, 78–80, 157, 269, 282, 300, 312
 Sigurðsson, Sigurður 61, 63, 158, 286
 Silbersweig, David A. 46, 304
 Silverman, B. W. 62, 310
 Simmen, Martin W. 95, 284
 Simmons, A. 79, 274
 Simon, Tony J. 21, 311
 Singer, Michael S. 93, 310
 Singpiel, F. 30, 55–57, 312
 Sinnweill, Teresa M. 46, 311
 Sittenfeld, M. 31, 307
 Sjølund, Arne 87, 311
 Skerjanc, R. 88, 311
 Skidmore, B. 98, 277
 Skinhøj, Erik 24, 295
 Skoufos, Emmananouil 93, 310
 Skudlarski, Pawel 78, 79, 123, 299, 310
 Sled, J. G. 42, 94, 269, 276, 309
 Small, Henry G. 87, 99, 102, 105, 311
 Smart, John C. 105, 269
 Smith, A. F. M. 37, 296
 Smith, Anne M. 46, 311
 Smith, C. W. 88, 278
 Smith, Carolyn Beebe 41, 314
 Smith, Clayton 31, 283
 Smith, Edward E. 125, 269
 Smith, J. R. 31, 307
 Smith, John R. 96, 311
 Smith, S. 45, 318
 Smith, Stephen 42, 43, 93, 289, 311
 Smith, Stephen M. 46, 270
 Smyth, Padhraic 39, 51, 311
 Snyder, Abraham Z. 95, 305
 Snyder, Herbert 96, 311
 Snyder, Timothy Law 103, 311
 Soboroff, Ian M. 100, 311
 Söderberg, Bo 70, 303
 Sokoloff, Louis 25, 41, 298, 314
 Solla, Sara A. 37, 70, 295
 Sommer, Friedrich T. 52, 269
 Sommer, Ove 85, 311
 Somorjai, Ray L. 46, 48, 52, 93, 269, 305, 309, 311
 Sonka, Milan 81, 311
 Sorensen, J. A. 46, 48, 276, 297
 Sørensen, N. L. 88, 286
 Sørensen, S. E. B. 88, 286
 Sparck Jones, Karen 102, 308, 311
 Specht, Donald F. 60, 138, 311
 Spelbring, D. R. 43, 304
 Spelke, Elizabeth 21, 22, 277, 312
 Sperling, C. D. 30, 285
 Spiegel, Murray R. 151, 152, 312
 Spire, Jean-Paul 28, 275, 314
 Stamatakis, E. A. 43, 312
 Stanescu, R. 21, 277
 Stata, Raymie 104, 273
 Steele, J. Michael 103, 311
 Steenblik, Richard A. 88, 312
 Steensgaard, P. L. 26, 271
 Steinberg, Tilmann 97, 279
 Steinmetz, H. 27, 315
 Stephan, K. E. 95, 303
 Stephan, Klaas E. 94, 95, 312
 Stephenson, Ken 90, 289
 Stern, Emily 46, 304
 Stern, J. J. 42, 269, 309
 Stern, P. 87, 305
 Stetter, Martin 30, 55–57, 309, 310, 312
 Stevenson, J. 123, 310
 Stiefel, E. 36, 287
 Stock, D. 37, 70, 287
 Stock, David G. 37, 287
 Stødkilde-Jørgensen, H. 46, 305
 Stone, James Verne 57, 312
 Stone, M. 39, 312
 Stone, Mervyn 74, 273
 Stork, David G. 61, 278
 Streeter, Lynn Q. 102, 277
 Strick, P. L. 98, 112, 305
 Strother, Stephen C. 18, 39, 42–45, 48, 49, 51, 52, 54, 55, 70, 71, 75, 78–80, 90, 91, 93, 117, 145, 157, 158, 188–221, 266, 269, 282, 286, 291, 293–295, 300, 301, 303, 307, 312, 313
 Strupp, John P. 26, 93, 300, 313
 Studholme, C. 43, 315

- Suetens, P. 43, 298, 315
 Sugiura, Motoaki 123, 289
 Sum, J. M. 17, 277
 Sumanaweera, T. S. 43, 315
 Summerlin, J. L. 24, 81, 95, 294
 Summers, R. 52, 269
 Sumners, De Witt L. 90, 91, 289, 307
 Sundin, A. 43, 268
 Svarer, Claus 38, 39, 42, 50–52, 54, 55, 65, 68, 70, 71, 84, 117, 127, 128, 157, 159, 189–199, 269, 285, 286, 294, 295, 300, 301, 303, 313
 Svensson, P. 44, 94, 96, 281, 308
 Swithenby, S. J. 28, 287
 Switzer, P. 56, 313
 Syed, I. 28, 314
 Szikla, G. 109, 124, 313
 Szilagyi, Peter 107, 315

 Takashima, S. 31, 290
 Takashima, Sachio 31, 290
 Takazawa, K. 52, 314
 Takechi, O. 79, 269
 Takei||, Naoyuki 30, 293
 Talairach, J. 109, 294
 Talairach, Jean 23, 24, 44, 81, 84, 90, 94, 95, 109, 124, 313
 Tamiillo, Brenden 140, 294
 Tank, D. W. 30, 303
 Tapanainen, Pasi 100, 285
 Tarassenko, Lionel 62, 63, 308, 313
 Taub, Edward 18, 293
 Taylor, Chris 28, 52, 269
 Taylor, David Scot 105, 278
 Taylor, J. G. 48, 288
 Teller, A. H. 137, 300
 Teller, E. 137, 300
 Ter-Pogossian, M. M. 28, 290
 Ter-Pogossian, Michael M. 29, 313
 Teukolsky, Saul A. 36, 306
 Thayer, Dorothy T. 54, 309
 Thesen, Stefan 43, 278
 Thiel, A. 90, 287
 Thiel, Thorsten 30, 298
 Thierry, G. 68, 313
 Thomas, M. R. 75, 293
 Thompson, P. D. 123, 316
 Thompson, W. T. 125, 292
 Thompson, William L. 123, 292
 Thorndike, R. L. 51, 313
 Thurfjell, Lennart 44, 289
 Tian, B. 129, 315
 Tiao, George C. 66, 69, 272
 Tibshirani, Robert J. 37, 60, 69, 79, 136, 287, 313
 Tikhonov, A. N. 37, 313
 Tipping, Michael E. 53, 54, 139, 313
 Tishby, Naftali 27, 307
 Tkach, J. 52, 278
 Toft, Peter 18, 42, 48, 51, 52, 54, 57, 60, 65, 68, 93, 103, 157–159, 175–187, 266, 285, 286, 303, 313
 Toft, Peter Bjerre 17, 26, 272
 Toga, Arthur W. 44, 95, 294, 299, 316
 Tomkins, Andrew 96, 101, 104, 273, 274
 Toni, I 143, 308
 Toronov, Vlad 31, 280
 Torrey, H. C. 29, 306
 Tournoux, Pierre 23, 24, 44, 81, 84, 90, 94, 95, 109, 124, 313
 Towle, Vernon L. 28, 275, 314

 Toyama, H. 52, 314
 Travère, J. M. 114, 315
 Truwit, C. L. 17, 296
 Tsai, C.-L. 40, 289
 Tsaparas, Penayiotis 105, 272
 Tsivkin, S. 21, 277
 Tuch, David S. 30, 51, 52, 314, 316
 Tukey, J. W. 99, 102, 276
 Tulving, E. T. 79, 80, 299
 Tulving, Endel 98, 123, 296, 314
 Turk, Greg 86, 303
 Turkeltaub, Peter 62, 98, 126, 314
 Turkheimer, Federico 41, 314
 Turkington, T. G. 43, 314
 Turnbull, L. W. 71, 292
 Turner, R. 30, 65, 281, 293
 Turner, Robert 26, 30, 40, 41, 47, 48, 54, 55, 65, 71, 79, 80, 277, 281, 282, 290, 314
 Turski, P. A. 46, 276
 Tyler, Lorraine 30, 46, 277
 Tyler, Lorraine K. 30, 46, 277
 Tzourio, Christophe 44, 306

 Uemura, K. 52, 117, 295, 314
 Ugurbil, Kamil 26, 30, 300, 303, 317
 Ullman, Jeffrey D. 100, 267
 Ulmer, John L. 60, 271
 Ungerleider, Leslie G. 79, 90, 95, 96, 129, 287–289, 298, 300
 Uselton, S. P. 86, 282
 Uylings, H. 44, 94, 96, 281, 308
 Uylings, Harry B. 23, 24, 268

 v. Keyserlingk, D. Graf 90, 270
 Vaillant, M. 94, 273
 Valind, S. 43, 268
 van den Elsen, P. A. 43, 315
 Van den Wollenberg, A. L. 74, 314
 van Erp, T. G. 25, 300
 Van Essen, David C. 45, 81, 84, 89, 90, 278, 280, 314
 Van Gelder, Allen 85, 314, 316
 Van Horn, John D. 94, 96, 285
 Van Lare, J. 129, 315
 Van Loan, Charles F. 136, 284
 Vandenbergh, R. 99, 306
 Vandermeulen, D. 43, 298, 315
 VanMeter, J. 129, 315
 Vanzetta, Ivo 26, 31, 314
 Vasileios Megalooikonomou, Christos Davatzikos 99, 314
 Vazquez, Alberto L. 26, 45, 303, 314
 Velez, Bienvenido 107, 315
 Veltman, D. J. 30, 125, 306, 314
 Venugopal, R. 45, 284
 Vérard, Laurent 114, 315
 Verwichte, Erwin 88, 315
 Vetterling, William T. 36, 306
 Vevea, J. M. 30, 270
 Viergever, M. A. 43, 315
 Viergever, Max A. 41, 298
 Villes, Charles L. 99, 104, 281
 Villringer, A. 31, 302
 Villringer, Arno 30, 315
 Vinod, H. 50, 315
 Vinod, H. D. 74, 315
 Viola, Paul 102, 289
 Vivanco, R. 52, 311
 Vivanco, Rodrigo A. 93, 305

- Viveros, Roman 66, 71, 74, 76, 149, 273
 Vo, K. P. 87, 283
 Vogelmann, Volker 87, 291
 Volbracht, Sabine 88, 315
 Volkow, Nora 90, 301
 von Cramon, D. Y. 48, 277
 von Cramon, D. Yves 17, 41, 65, 69, 93, 289, 292, 297, 307
 von Economo, Constantin 23, 315
 von Stockhausen, H.-M. 90, 287
- Wada, J. 17, 315
 Wagner, A. D. 17, 277
 Wagner, Robert A. 99, 315
 Wahren, Waldemar 90, 309
 Walker, S. 102, 308
 Walter, H. 27, 315
 Wang, M. Y. 43, 315
 Ward, B. Douglas 41, 315
 Watabe, Hiroshi 90, 315
 Watanabe, E. 31, 298
 Watson, J. D. G. 44, 116, 317
 Weaver, J. B. 55, 315
 Wedeen, Van J. 30, 51, 52, 314–316
 Wei, T. C. 143, 287
 Weichselbaum, R. R. 43, 304
 Weiller, C. 122, 125, 290, 315
 Weiller, Cornelius 123, 288
 Weiskoff, Robert M. 30, 66, 303
 Weiss, Ron 107, 315
 Weiss, W.-D. 90, 287
 Weisskoff, R. M. 30, 270, 293
 Weisskoff, Robert M. 46, 47, 306
 Wells, William M. 90, 283
 Welsh, Tom 90, 301
 Welsh, Tomihisa 90, 318
 Wennerås, Christine 120, 315
 Wentz, K. U. 89, 315
 Wenzel, R. R. 31, 302
 Wessinger, C. M. 129, 315
 West, Ben 84, 278
 West, Jay 43, 315
 West, R. 79, 297
 Westad, Frank 77, 288
 Westerfield, Marissa 28, 298
 Westermann, Rüdiger 85, 311
 Wexelblat, Alan 87, 316
 Whalen, P. J. 143, 274
 Wheless, J. W. 31, 307
 White, Howard D. 19, 99, 104, 105, 135, 297, 316
 White, Thomas 94, 283
 Whitehead, Sara H. 94, 296
 Wählin, J. 44, 308
 Widman, Guido 28, 316
 Widmayer, Peter 105, 278
 Wiegell, Mette Regin 30, 51, 52, 316
 Wiener, Janet 104, 273
 Wienhard, K. 90, 287
 Wiesel, T. N. 30, 285
 Wiggs, C. L. 129, 298
 Wilbur, W. J. 102, 316
 Wildgruber, Dirk 52, 269
 Wildschütz, G. 26, 271
 Wilhelms, Jane 85, 316
 Williams, Chris K. I. 69, 316
 Williams, E. J. 46, 302
 Williams, P. M. 37, 316
- Williams, S. 43, 282
 Williams, S. C. 78–80, 273
 Williams, S. C. R. 40, 47, 79, 274, 282
 Williams, Steve C. R. 65, 273
 Wills, Adrian J. 123, 316
 Willshaw, David J. 95, 284
 Wilson, J. T. L. 43, 312
 Wilson, James 30, 46, 277
 Wilson, Orion 85, 316
 Wise, R. 125, 316
 Wise, R. J. S. 71, 273
 Wise, Richard 123, 288
 Witten, Ian H. 96, 302
 Witzel, T. 95, 303
 Wold, Agnes 120, 315
 Wold, H. 73, 316
 Wold, Herman 73, 138, 316
 Wold, S. 73, 78, 316
 Woldorff, M. 17, 317
 Woldorff, Marty 95, 294
 Woldorff, Marty G. 94, 95, 294
 Wolf, Martin 31, 280
 Wolfe, P. 36, 316
 Wolff, G. 37, 70, 287
 Wolkow, Nora 90, 318
 Wolpert, D. H. 37, 316
 Wolpert, Robert L. 139, 270
 Wong, A. 100, 309
 Wong, Eric C. 25, 26, 46, 66–68, 71, 108, 269, 274
 Woodruff, P. W. R. 79, 274
 Woods, R. P. 125, 315
 Woods, Roger P. 43, 44, 55, 93, 116, 123, 285, 312, 315–317
 Woodward, Jeffrey B. 94, 96, 285
 Woolrich, Mark 93, 311
 Wöpkling, Matthias. 88, 304
 Worden, K. 63, 317
 Wormell, Irene 115, 317
 Worsley, Keith J. 17, 39–41, 47, 48, 68, 69, 71, 72, 74, 93, 118, 269, 274, 282, 296, 306, 317
 Worth, A. J. 42, 269, 309
 Wowk, B. 46, 52, 309
 Wright, I. 79, 274
 Wyper, D. J. 43, 312
- Xiong, J. 17, 41, 317
 Xiong, Jin-Hu 62, 98, 126, 136, 280
 Xiong, Jinhu 49, 317
 Xu, X. L. 43, 312
- Yacoub, Essa 26, 317
 Yacoup, Essa 26, 317
 Yagel, Roni 85, 310
 Yamane, Yohei 87, 99, 290
 Yamashita, Y. 31, 298
 Yan, Yihong 46, 304
 Yanaka, Kiyoyuki 30, 293
 Yang, C. S. 100, 101, 309
 Yang, Y. 102, 316
 Yang, Yihong 46, 311
 Ye, Frank Q. 46, 311
 Yee, Thomas W. 71, 317
 Yoshimura, Takashi 51, 267
 Yoshizawa, Shuji 40, 301
 Young, G. 53, 278
 Young, M. P. 95, 318
 Young, Malcolm P. 94, 95, 312

- Young, Peter 83, 318
Ypma, Alexander 62, 63, 318

Zamarippa, Frank 62, 280
Zamarripa, Frank 62, 98, 126, 136, 280
Zarahn, Eric 25, 45–48, 103, 108, 267, 282, 318
Zatorre, R. J. 84, 122, 123, 295, 318
Zavaljevski, A. 45, 318
Zeffiro, Thomas A. 62, 98, 126, 314
Zeger, Scott L. 47, 65, 69, 294
Zeien, Gene 93, 284

Zeleznik, Michael P. 86, 318
Zhang, Yongyue 45, 93, 311, 318
Zhao, B. 89, 315
Zhu, Wei 90, 301, 318
Ziljenbos, A. P. 94, 276
Zilles, Karl 23, 24, 43, 44, 94, 96, 268, 281, 283, 290, 301, 308, 310, 318
Zipf, George Kingsley 101, 318
Zuck, E. 129, 310
Zysset, S. 69, 292

Index

- abundance matrix, 103, 117, **135**
- ACA, **140**
- ACF, **140**
- acknowledgment, 155
- activation function, **135**
- active learning, **135**
- adaptive principal components extraction, **135**
- adjacency matrix, 102, 103
- agnostic labeling, 79
- AIC, 39, 40, 51, 61, 131, **140**
- AIR, 43, 44, 93, 116, **140**
- alignment
 - intersubject, 43
 - intrasubject, 43
- aMRI, **140**
- analysis of covariance, **135**
- ANCOVA, 39, 48, 65, 66, **140**
- ANIMAL, 45, 95
- ANOVA, 65, **140**
- anterior, **142**
- anti-Hebbian learning, **135**
- APEX, **140**
- AR, **140**
- ARMA, **140**
- ARMAX, **140**
- ARMS, **140**
- ARS, **140**
- ASCII, **140**
- asymmetric divergence, **135**
- author cocitation analysis, 99, 105, 108, **135**
- authority, 104
- auto-association, **135**
- autocorrelation, 46
-
- BA, **140**
- backpropagation, **135**
- bag-of-words, 100, 134
- behavioral component, 21
- BGO, **140**
- bias, 35, **135**
- bias-variance trade-off, **135**
- BIC, 39, **140**
- bilinear model, 64, 76, 77
- bioinformatics, 93
- BOLD, 30, **140**
- Bq, **140**
- BRAID, 94, 99
- BrainMap, 3, 5, 19, 21, 22, 27, 28, 81, 91, 94–96, 98, 120–129, 131, 132, 144, 158
- BrainWeb, 44, 94
- Broca's area, 93
-
- Brodmann
 - area, 23, 24, 143
 - number, 81, 84, 94, 95, 143
- bullseye visualization, 87, 112, 113
- Burt matrix, 53, **135**
-
- canonical
 - analyses, 71
 - correlation analysis, 71
 - correlation variables, 72
 - correlation vectors, 72
 - discriminant analysis, 72
 - ridge analysis, 74
 - variables analyses, 74
 - variate analysis, 79
 - variate analysis, 54, 72
- canonical correlation analysis, **135**
- canonical variable analysis, **135**
- canonical variate analysis, **135**
- caudal, **142**
- CBF, 25, **140**
- CBV, 25
- CCA, **140**
- centering matrix, 133
- CGI, **140**
- chi-square metric, 53
- Ci, **140**
- cluster, **135**
- cluster analysis, 50
- coauthor analysis, 112
- cocitation analysis
 - author, 105, 108, 267
 - journals, 114
- COG, **141**
- cognition, **135**
- cognitive component, 21
 - episodic memory retrieval, 79
 - face processing, 79
- colored noise, 40
- conditional (differential) entropy, **135**
- confound, 45, **135**
- conjugate prior, 66, **135**
- connected component, 81, 104, 105, 113
- consistent, **135**
- continuum regression, 74
- contralateral, **142**
- coregistration, 43
- Corner Cube environment, 90, 91, 121, 122, 125, 126
 - on the Web, 94
- correspondence analysis, 52, 53
- cost function, 34, 133, **135**

- cross-correlation, 67
- cross-correlation asymmetric principal component analysis, 73
- cross-entropy, **135**
- cross-validation, **135**
- CSF, **141**
- CT, **141**
- CVA, **141**

- DAG, **141**
- data matrix, 33, 102, 134
- data mining, 97
- DBM, **141**
- dependence, **136**
- dependent variable, **136**
- derivations, 145
- design matrix, 65, **136**
- diameter, 104
- differential entropy, **136**
- digraph, 102
- directed divergence, **136**
- distance matrix, 102
- DOF, **141**
- dorsal, **142**
- DTI, **141**
- DWI, 30, **141**

- ED, **141**
- EEG, 27, **141**
- efMRI, **141**
- eigenauthor, 109
- eigenvalue decomposition, 56, 72, 76, 145–148, 150
- elliptic distribution / elliptically contour distribution, **136**
- EM, 54, 61, **141**
- EMG, **141**
- empirical Bayes, **136**
- ENG, **141**
- Entrez, 96
- entropy, **136**
 - differential, 151
 - Kullback-Leibler, 151
- EOG, **141**
- EPI, **141**
- erfMRI, **141**
- ERP, 27
- ESM, **141**
- estimation, 34, **136**
- evidence, 38, **136**
- expectation maximization, **136**
- explorative statistics, **136**

- F-masking, 48, **136**
- FA, 30, **141**
- factor analysis, 54, 64, 77
- FAIR, **141**
- FDG, **141**
- feed-forward neural network, **136**
- FID, 29, **141**
- file drawer problem, 97
- finite impulse response (model), **136**
- FIR, 65, **141**
- fMRI, 17, 30, **141**
- fMRIDC, 129
- FOV, **141**
- FPE, 40
- Frobenius norm, 133, **136**
- full width half maximum, **136**
- functional integration, **136**
- functional segregation, **136**
- functional volumes modeling, 98, 126, **136**
- fundus, **142**
- FVM, 98, **141**
- FWHM, **141**

- Gauss-Newton (method), **136**
- GE, **141**
- general linear model, 65, **136**
- generalized additive models, **136**
- generalized inverse, 75, **136**
- generalized least squares, **137**
- generalized singular value decomposition, 74
- GLM, 46, 65, 69, **141**
- GLMM, **141**
- GLS, **141**
- glyph, 86
- Good-Turing frequency estimation, **137**
- Google™, 96, 100, 101, 105, 107
- Graphviz, 87

- H2-15O, **141**
- headset, 88
- Hebbian learning, **137**
- hemodynamic response
 - negative, 26
 - nonlinear, 26
- hemodynamic response function, **137**
- Hessian, 133
- heteroassociation, **137**
- heteroencoder, 74
- Hinton diagram, 87
- HME, **141**
- HPD, **141**
- HRF, **141**
- HTML, **141**
- HTTP, **141**
- hub, 104
- hyperparameter, **137**

- ICA, 46, 56, 57, **141**
- ICBM, **141**
- identification, 34, **137**
- idf, 101, 102, 111, 116, **141**
- IIR, **141**
- ill-posed, **137**
- Imax, **137**
- independent component analysis, 56
 - spatial, 57
 - temporal, 57
- indicator matrix, 103
- inferior, **142**
- information retrieval, 99
- inhomogeneity correction, 42

- inion, **137**
- Inquirus, 104
- International Consortium for Brain Mapping, **137**
- inverse document frequency, 102, 111, 116
- inversion, 34, **137**
- iOIS, **141**
- ipsilateral, **142**
- IR, **141**
- IRF, **141**
- ISI, **141**
- ISI Web of Science, 96, 101, 105, 106, 108, 109, 111, 127
- ITI, **141**
- K-means, 50, **137**
- k-nearest neighbor, **137**
- Karhunen-Lóeve transformation, **137**
- kernel density estimation, 98, **137**
- kernel methods, 61
- Kullback-Leibler distance, **137**
- lateral, **142**
- lateral orthogonalization, 73, 74, **137**
- learning, 34, **137**
- leave-one-out, 39, 62, **137**
- LED, **141**
- likelihood, 34, 133, **137**
- link analysis, 102
- link function, **137**
- lix, 101, **137**
- LMS, **141**
- log-likelihood, 133
- LON, **141**
- LOO, **141**
- Lyngby, 18, 37, 38, 51
- m-estimation, **137**
- MAF, 56
- MANCOVA, **141**
- manifold, **137**
- MANOVA, **141**
- MAP, 34, **141**
- Marching Cube, 85
- Markov chain Monte Carlo, **137**
- mass-univariate, 65
- mass-univariate statistics, **137**
- matrix
 - adjacency, 102
 - data matrix, 102
 - topology, 102
- maximal eigenvector, **137**
- maximum a posteriori, 34, 70, **137**
- maximum likelihood, 34, **137**
- MCMC, **141**
- MDL, **141**
- MDS, **141**
- medial, **142**
- MEDLINE, 96, 114
- MEG, 27, **141**
- meta-analysis, 97
- Metropolis-Hasting algorithm, **137**
- mixture model, 60
- MLE, **141**
- Molgedey-Schuster, 56
- Moore-Penrose inverse, 75, **137**
- motion correction, 43
- MRF, **141**
- MRI, 29, **142**
 - diffusion, 30
 - perfusion, 30
- MRS, **142**
- multidimensional scaling, 54
- multiple regression analysis, **137**
- multiple regression model, 65
- multivariate analysis, **137**
- multivariate regression analysis, **138**
- multivariate regression model, **138**
- mutual information, 43, **138**, 148
- N-gram, 100
- nasion, **138**
- NCS, **141**
- network visualization, 87
- neural network, 70, **138**
- NeuroImage, 107, 114
- neuroinformatics, 18, 93
 - tools, 93
- neurological convention, 90, **138**
- NeuroNames, 23, 87, 94, 143
- NIRS, 17
- NMR, **141**
- non-informative prior, **138**
- non-negative matrix factorization, 55, 111
- non-parametric (model/modeling), **138**
- nonlinear modeling, 69
- normalization, 47
- novelty, 62, **138**
- novelty detection, 62, 120
- Nuages, 86, 125, 126
- nuisance, **138**
- object, **138**
- OIS, **141**
- optimal brain damage, 37, 70
- optimal brain surgeon, 37, 70
- optimization, 34, 35, **138**
- ordination, **138**
- orthogonal, **138**
- orthonormal, **138**
- outlier detection, 62, 120
- outliers, 120
- PACF, **141**
- PageRank™, 105
- partial least squares, 72, 73, 79
 - orthonormalized, 74
- partial least squares (regression), **138**
- Parzen window, **138**
- Pathfinder, 87
- PCA, **141**
- PCoA, **141**
- PCR, 66, **141**
- PDA, 75

- PDF, **142**
pdf, **141**
penalized discriminant analysis, 75, **138**
perceptron, **138**
periauricular, **138**
PET, 17, 28, **142**
PLS, **142**
PLSA, **142**
pMRI, **142**
PMT, **142**
polysemy, **138**
posterior, 35, **142**
PPCA, 75, **142**
prediction, **138**
preliminary principal component analysis, **138**
preprocessing, 41
principal component analysis, 52, **138**
 preliminary, 79
 probabilistic, 54
 sensible, 54
principal component regression, **138**
principal coordinate analysis, **138**
principal covariate regression, **138**
principal manifold, **138**
prior, 37
probabilistic neural network, **138**
probabilistic principal component analysis, **139**
Probability density estimation, 60
probability density estimation
 kernel method, 61
profile likelihood, **139**
PSTH, **142**
Pubmed, 19, 96, 100–102, 108, 114, 116, 121, 122, 127, 129
 identifier, 96

radiological convention, 90, **139**
rank, 74, **139**, 140, 145, 146
rank deficient, **139**
ranking, 104
rCBF, 94, **142**
rCBV, **142**
rCMR, **142**
rCMRglu, **142**
rCMRO2, **142**
rcounts, **142**
realignment, 43
reconstruction, 42, 85
redundancy, 120
redundancy analysis, 74
region identification, 105
registration
 cross-modality, 43
 intermodality, 43
regularization, 37, **139**
related authors, 116
relative entropy, **139**
relCBF, **142**
ResearchIndex, 19, 96, 99, 101, 104, 106, 108, 119
RF, **142**

ridge
 canonical ridge analysis, 74
 regression, 37, 65, 66
robust statistics, **139**
rOEF, **142**
ROI, **142**
rostral, **142**
run, **139**

S/N, **142**
saliency map, 41, 70, **139**
saturation recovery, **139**
scientography, 87
SE, **142**
search, 34
self-organizing learning, **139**
sensible principal component analysis, **139**
session, **139**
sICA, **142**
sigmoidal, 70, **139**
similarity matrix, 102
singular value decomposition, 53, 54, 56, 66, 72–75, 104, 107, 109, 115, 145, 147, 150, 267
 generalized, 74
slice timing correction, 42, **139**
smoothing
 temporal, 46
SNR, **142**
SOA, **142**
sociogram, 87
softmax, 71, **139**
SOM, **142**
sparse coding, 55
spatial filtering, 48
spatial independent component analysis, **139**
spatial normalization, 43
SPCA, **142**
SPI, **142**
SPM, 25, 41, 43, **142**
SR, **142**
state-space, 79
statistic, **139**
statistical parametric images, **139**
statistical parametric mapping, **139**
statistical parametric maps, **139**
stemming, 101
stop words, 109
stripping, 42
superior, **142**
supervised (learning/pattern recognition), **139**
Sv, **142**
SVC, **142**
SVD, 53, **142**
synonymy, **139**
system, **139**
system identification, **139**

TAC, **142**
Talairach
 atlas, 24, 84, 86, 95
 coordinate, 94, 95, 120

Daemon, 24, 45, 81, 94, 95, 125, 127
TE, 30, **142**
temporal independent component analysis, **140**
term, 100
term analysis, 100
term identification, 100
test set, 34, 39, **140**
testing, 38
text analysis, 99
TI, 30
tICA, **142**
time-activity curve, **140**
TLS, **142**
tokenization, 100, 108
total least square, **140**
TR, 30, **142**
training, 34, **140**
training set, 34, 39, **140**
tree visualization, 87
trial, **140**
truncated singular valued decomposition, **140**
TSVD, **142**

univariate statistics, **140**
unsupervised learning, **140**
URL, **142**

VAC, **142**
validation set, **140**
VBM, **142**
ventral, **142**
Virtual Reality Modeling Language, 84
visualization
 bullseye, 112, 113
 graph, 87
 information, 91
 network, 87
 scientific, 89
VOI, **142**
volume rendering, 85
voxel, **140**
voxelization, 86
VPC, **142**
VRML, 29, 84, 91, 127, 128, **142**, 267

Wada test, 17
web
 search engine, 96
weight decay, 37, 70
weights, **140**
white noise, 40, **140**
word stemming, 101

z-score, **140**